



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/12, 15/54, 9/12 // C12Q 1/68, C12N 15/11	A1	(11) International Publication Number: WO 93/15201 (43) International Publication Date: 5 August 1993 (05.08.93)
(21) International Application Number: PCT/US93/00586 (22) International Filing Date: 22 January 1993 (22.01.93) (30) Priority data: 826,935 22 January 1992 (22.01.92) US (60) Parent Application or Grant (63) Related by Continuation US 826,935 (CIP) Filed on 22 January 1992 (22.01.92) (71) Applicant (for all designated States except US): NEW ENGLAND DEACONESS HOSPITAL [US/US]; 185 Pilgrim Road, Boston, MA 02215 (US).	(72) Inventors; and (75) Inventors/Applicants (for US only) : AVRAHAM, Hava [IL/US]; 50 Radmor Road, Brighton, MA 02135 (US). GROOPMAN, Jerome [US/US]; 79 Druce Street, Brookline, MA 02146 (US). COWLEY, Sally [GB/GB]; 28 Avonmore Road, London W14 8RS (GB). SCADDEN, David [US/US]; 62 Lexington Street, Weston, MA 02193 (US). (74) Agents: GRANAHAH, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US). (81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(54) Title: NOVEL PROTEIN TYROSINE KINASES		
(57) Abstract <p>The identification and isolation of novel protein tyrosine kinase genes present on human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LJ	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LJ	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

NOVEL PROTEIN TYROSINE KINASESDescriptionBackground of the Invention

Transduction of signals that regulate cell growth and differentiation is regulated in part by phosphorylation of various cellular proteins. Protein tyrosine kinases are enzymes that catalyze this process. Moreover, many act as growth factor receptors.

Summary of the Invention

10 The present invention relates to novel protein tyrosine kinase genes present in human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.

The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. The pTK genes have been shown to be present in both megakaryocytic and lymphocytic cells.

The pTK genes of the present invention show significant sequence homology with members of the c-kit subgroup of growth factor receptors with protein tyrosine kinase activity. The c-kit subgroup of receptor tyrosine kinases catalyze the phosphorylation of exogenous substrates, as well as tyrosine residues within their own polypeptide chains. (Ullrich, A. and Schlessinger, J., Cell, 61:203 (1990)). Members of the c-kit subgroup include FLT/FLK (Fetal Liver Kinase), FGF (Fibroblast

-2-

Growth Factor Receptor) and NGF (Nerve Growth Factor Receptor).

In particular, fourteen pTK genes have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor tyrosine kinase) were identified in megakaryocytic cells. Five pTK genes, referred to as LpTKs, were identified in lymphocytic cells and have been shown to be present in megakaryocytes as well. One pTK gene, referred to as HpTKs, was identified in human hepatoma cells. Six pTK genes, referred to as bpTK genes, found in human brain tissue.

SAL-S1 is related to the FLT/FLK family of pTKs. SAL-D4 is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3) is related to the NGF receptor family of pTKs.

The pTK genes, which are the subject of the present invention, were identified using two sets of degenerative oligonucleotide primers: a first set which amplifies all pTK DNA segments (SEQ ID NOS:1-2), and a second set which amplifies highly conserved sequences present in the catalytic domain of the c-kit subgroup of pTKs (SEQ ID NOS:3-4). The pTK genes identified in this manner are described below.

SAL-S1 is expressed in several megakaryocytic cell lines, but not in erythroid cell lines. The nucleotide sequence of SAL-S1 was obtained, revealing a sequence containing 158 base pairs. (SEQ ID NO:5). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:6) which exhibited significant sequence homology with known protein tyrosine kinases of the FLT/FLK family. The full

-3-

length gene sequence (SEQ ID NO: 17) contains 6827 b.p. and the deduced amino acid sequence (SEQ ID NO: 18) contains 349 residues.

SAL-D4, also expressed in megakaryocytic cells, is a DNA fragment containing the nucleotide sequence of 141 base pairs. (SEQ ID NO:7). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:8) which exhibited significant sequence homology with known protein tyrosine kinases of the FGF receptor family.

10 The LpTKs, including LpTK 2, LpTK 3, LpTK 4, and LpTK 13 and LpTK 25, are expressed in lymphocytic cells, as well as megakaryocytic cells. The nucleotide sequence (151 base pairs) of the LpTK 3 gene was obtained (SEQ ID NO:11), and exhibited significant homology with known
15 protein tyrosine kinases of the NGF receptor family. The nucleotide sequences of the LpTK 2, LpTK 4, and LpTK 13 genes contained 149 base pairs (SEQ ID NO:9), 137 base pairs (SEQ ID NO:13), and 211 base pairs (SEQ ID NO:15) respectively. LpTK 25 has a nucleotide sequence of 3120
20 b.p. (SEQ ID NO: 22). A full length gene sequence has been obtained for LpTK 2 (SEQ ID NO: 19) which contains 7606 b.p. Additional sequencing of LpTK 4 revealed a sequence of 404 b.p. (SEQ ID NO: 21).

The HpTK 5 gene, expressed in human hepatoma cells,
25 has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). Nucleotide sequences of the bpTK's, including bpTK 1, bpTK 2, bpTK 3, bpTK 4, bpTK 5 and bpTK 7 are expressed in human brain tissue encode proteins having the amino acid sequences of SEQ ID NOS: 25-30 respectively.

30 Thus the present invention includes DNA isolated from a human megakaryocytic cell line, which hybridizes to a

-4-

DNA fragment which hybridizes to DNA encoding an amino acid sequence which is highly conserved in the catalytic domain of protein tyrosine kinases of the c-kit subgroup.

The present invention also includes the proteins
5 encoded by the pTK genes identified as described herein, which exhibit significant sequence homology with members of the c-kit subgroup of pTKs (i.e. FLT/FLK (SAL-S1), FGF receptor (SAL-D4) or NGF receptor (LpTKS)) as well as the proteins encoded by HpTK 5 and the bpTKs. The present
10 invention also includes SAL-S1, SAL-D4, and LpTK, HpTK and bpTK homologues or equivalents (i.e., proteins which have amino acid sequences substantially similar, but not identical, to that of SAL-S1, SAL-D4, the LpTKs HpTK and the bpTKs, which exhibit tyrosine kinase activity.) This
15 invention further includes peptides (SAL-S1, SAL-D4, LpTK, HpTK and bpTK fragments) which retain tyrosine kinase activity, yet are less than the entire SAL-S1, SAL-D4, LpTK, HpTK and bpTK sequences), monoclonal and polyclonal antibodies specific for SAL-S1, SAL-D4, the LpTKs, HpTK
20 and the bpTKs, and uses for the SAL-S1, SAL-D4, the LpTK, HpTK and the bpTK nucleic acid sequences and SAL-S1, SAL-D4, LpTK, HpTK and bpTK equivalents.

The present invention further includes nucleic acid sequences which hybridize with DNA or RNA encoding the
25 proteins described herein, which exhibit significant sequence homology with the FLT/FLK, FGF receptor or NGF receptor family of protein tyrosine kinases contained within the c-kit subgroup. Such nucleic acid sequences are useful as probes to identify pTK genes in other
30 vertebrates, particularly mammals, and in other cell types. They can also be used as anti-sense

-5-

oligonucleotides to inhibit protein tyrosine kinase activity, both in vitro and in vivo.

The SAL-S1, SAL-D4, LpTK, HpTK and bpTK, tyrosine kinases of the present invention can be used as target
5 proteins in conjunction with the development of drugs and therapeutics to modulate cell growth, differentiation and other metabolic functions. The SAL-S1, SAL-D4, LpTK, HpTK or bpTK proteins can be used as agonists or antagonists to other tyrosine kinases. The SAL-S1, SAL-D4, LpTK, HpTK or
10 bpTK tyrosine kinases can also be instrumental in the modulation of megakaryocyte and/or platelet adhesion interactions.

In addition, the SAL-S1, SAL-D4, LpTK, HpTK and bpTK tyrosine kinases can be used in screening assays to detect
15 cellular growth and/or differentiation factors. Using standard laboratory techniques, the ligands of the pTKs of the present invention can be identified. Once identified, assays can be designed to detect these ligands present endogenously, within cells, as well as exogenously, in
20 extra cellular fluids. Assays can also be designed as diagnostic aids to detect these ligands in body fluids such as blood and urine.

Brief Description of the Drawings

Figure 1 depicts the nucleotide sequence of SAL-S1
25 (SEQ ID NO: 5) and the deduced amino acid sequence (SEQ ID NO:6).

Figure 2 depicts the nucleotide sequence of SAL-D4 (SEQ ID NO:7) and its deduced amino acid sequence (SEQ ID NO:8).

-6-

Figure 3A depicts the nucleotide sequence (SEQ ID NO:9) and its deduced amino acid sequence (SEQ ID NO:10) for LpTK 2.

Figure 3B depicts the nucleotide sequence (SEQ ID NO:11) and its deduced amino acid sequence (SEQ ID NO:12) for LpTK 3.

Figure 3C depicts the nucleotide sequence (SEQ ID NO:13) and its deduced amino acid sequence (SEQ ID NO:14) for LpTK 4.

10 Figure 3D depicts the nucleotide sequence (SEQ ID NO:15) and its deduced amino acid sequence (SEQ ID NO:16) for the LpTK 13.

Figure 4A-4J depicts the full-length nucleotide sequence (SEQ ID NO: 17) and its deduced amino acid sequence (SEQ ID NO: 18) for SAL-S1.

Figure 5A-5J depicts the full length nucleotide sequence (SEQ ID NO: 19) and the deduced amino acid sequence (SEQ ID NO: 20) for LpTK2.

20 Figure 6 depicts the partial nucleotide sequence (SEQ ID NO: 21) for LpTK4.

Figure 7A-7D depicts the full length nucleotide sequence (SEQ ID NO: 22) for LpTK25.

25 Figure 8A-8F depicts the full length nucleotide sequence (SEQ ID NO: 23) and the deduced amino acid sequence (SEQ ID NO: 24) for HpTK5.

Figure 9 depicts the amino acid sequence (SEQ ID NO: 25) of bpTK1.

Figure 10 depicts the amino acid sequence (SEQ ID NO: 26) of bpTK2.

30 Figure 11 depicts the amino acid sequence (SEQ ID NO: 27) of bpTK3.

-7-

Figure 12 depicts the amino acid sequence (SEQ ID NO: 28) of bpTK4.

Figure 13 depicts the amino acid sequence (SEQ ID NO: 29) of bpTK5.

5 Figure 14 depicts the amino acid sequence (SEQ ID NO: 30) of bpTK7.

Detailed Description of the Invention

Novel protein tyrosine kinase genes have been identified, their nucleic acid sequences determined, and
10 the amino acid sequences of the encoded proteins deduced. The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology to
15 with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. These genes have been shown to be present in both megakaryocytic and lymphocytic cells.

To facilitate the isolation and identification of
20 these novel pTKs, two sets of DNA probes were used, as described in the Exemplification. The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2) (Matthews, W. Cell 65: 1143 (1991; Wilks, A. F. Proc. Natl. Acad. Sci. USA
25 86:1603 (1989)). These sequences were used as primers in a polymerase chain reaction to amplify tyrosine kinase DNA segments. (Mullis, K. et al., Cold Spring Harbor Symp. Advan. Biol. 51:263 (1986).

The second set consisted of two oligonucleotide
30 sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4)

-8-

designed to amplify the nucleic acid sequence which encodes the highly conserved regions of the catalytic domains of the c-kit family of protein tyrosine kinases. These sequences were used as primers in the polymerase chain reaction in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK primers were identified, isolated and subsequently sequenced.

In particular, fourteen pTK genes exhibiting significant homology with the c-kit subgroup of protein tyrosine kinases have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor) were identified in several megakaryocytic cell lines, including CMK 11-5, DAMI, UT-7 and UT-7 grown in erythropoietin, but not in the erythroid cell lines HEL, PMA stimulated HEL cells, or K562. Five pTK genes, referred to as LpTKs, were identified in lymphocytic, as well as in megakaryocytic cells. One pTK gene, referred to as HpTK5 was identified in human hepatoma cells and six genes, referred to as bpTKs, were identified in human brain tissue.

SAL-S1 (SEQ ID NO:6 and 18) encoded by the nucleic acid sequence SEQ ID NOS:5 and 17, exhibits significant homology with the FLT/FLK family of pTKs. SAL-D4 (SEQ ID NO:8) encoded by SEQ ID NO:7, is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3 (SEQ ID NO:12) encoded by the SEQ ID NO:11, is related to the NGF receptor family of pTKs. The remaining LpTKs, LpTK2 (SEQ ID NO:10) encoded by SEQ ID NO:9; LpTK4 (SEQ ID NO:14) encoded by SEQ ID NO:13; LpTK13 (SEQ ID NO:16) encoded by SEQ ID NO:15 LpTK25 encoded by SEQ ID NO: 22, also exhibit

-9-

sequence homology with known protein tyrosine kinases (Data not shown).

HpTK5 (SEQ ID NO: 24) encoded by SEQ ID NO: 23 and the bpTKs 1, 2, 3, 4, 5 and 7 (SEQ ID NOS: 25-30) respectively, also exhibit sequence homology with known protein tyrosine kinases.

Thus, as described above, DNA which hybridize with DNA encoding amino acid sequences present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases have been isolated and sequenced.

These isolated DNA sequences, collectively referred to as pTKs genes, (and their deduced amino acid sequences) have been shown to exhibit significant sequence homology with known members of receptor tyrosine kinase families.

Once isolated, these DNA fragments can be amplified using known standard techniques such as PCR. These amplified fragments can then be cloned into appropriate cloning vectors and their DNA sequences determined.

These DNA sequences can be excised from the cloning vectors, labeled with a radiolabeled nucleotide such as ³²P and used to screen appropriate cDNA libraries to obtain the full-length cDNA clone.

The pTk genes as described above have been isolated from the source in which they occur naturally, i.e. megakaryocyte and lymphocytic cells. The present invention is intended to include pTk genes produced using genetic engineering techniques, such as recombinant technology, as well as pTk genes that are synthesized chemically.

The deduced amino acid sequences of the pTK genes include amino acid sequences which encode peptides exhibiting significant homology with the catalytic domain

-10-

of protein tyrosine kinases of the c-kit subgroup of tyrosine kinases. These proteins, encoded by the pTk genes, can include sequences in which functionally equivalent amino acid residues are substituted for
5 residues within the sequence, resulting in a silent change, that is a change not detected phenotypically. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent,
10 resulting in a silent substitution.

In addition, the protein structure can be modified by deletions, additions, inversion, insertions or substitutions of one or more amino acid residues in the sequence which do not substantially detract from the
15 desired functional tyrosine kinases properties of the peptide.

Modified pTKs of the present invention, with receptor tyrosine kinase activity can be made using recombinant DNA techniques, such as excising it from a vector containing a
20 cDNA encoding such a protein, or by synthesizing DNA encoding the desired protein mechanically and/or chemically using known techniques.

An alternate approach to producing the pTKs of the present invention is to use peptide synthesis to make a
25 peptide or polypeptide having the amino acid sequence of such a protein. The peptides or modified equivalents thereof, can be synthesized directly by standard solid or liquid phase chemistries for peptide synthesis.

Preferably, the pTKs of the present invention will be
30 produced by inserting DNA encoding the proteins into an appropriate vector/host system where it will be expressed.

-11-

The DNA sequences can be obtained from sources in which they occur naturally, can be chemically synthesized or can be produced using standard recombinant technology.

This invention also pertains to an expression vector
5 comprising a pTK gene of the present invention, encoding for a protein which exhibits receptor tyrosine kinase activity.

The pTK genes of the present invention can be used for a number of diagnostic and therapeutic purposes. For
10 example, the nucleic acid sequences of the pTK genes can be used as probes to identify other protein tyrosine kinases present in other cell types, including eukaryotic and prokaryotic cell types.

The nucleic acid sequences can be used to design
15 drugs that directly inhibit the kinase activity of protein tyrosine kinases, or to design peptides that bind to the catalytic domain of tyrosine kinases, thus inhibiting their activity. These sequences can also be used to design anti-sense nucleotides that can also inhibit, or
20 destroy, tyrosine kinase activity. Such inhibition of tyrosine kinase activity would be desirable in pathological states where decreased cellular proliferation would be beneficial, such as leukemias or other malignancies.

25 The nucleic acid sequences can also be used to design drugs, peptides or anti-sense nucleotides as above, but with enhancing, rather than inhibitory effects, on tyrosine kinases. Such enhanced tyrosine kinase activity would result in increasing the phosphorylation of
30 substrates (exogenous, as well as endogenous tyrosine residues). Enhanced effects would be desirable in states

-12-

where increased cellular proliferation would be beneficial, such as anemias, bleeding disorders and during surgical procedures.

The pTK genes of the present invention can also be
5 used to obtain soluble fragments of receptor tyrosine kinases, capable of binding their respective ligands (i.e. fibroblast growth factor).

pTK genes encoding soluble receptor tyrosine kinase fragments can be produced using recombinant DNA techniques
10 or synthetically. In either case, the DNA obtained encodes a soluble pTK fragment which lacks a substantial portion of the hydrophobic transmembrane region to permit solubilization of the fragment.

These soluble pTK protein fragments can be introduced
15 exogenously to act as competitors with the endogenous, membrane bound pTK for their respective ligands, thus inhibiting tyrosine kinase activity. Alternately, a modified soluble pTK protein fragment can be introduced which binds the ligand but does not activate kinase
20 activity.

These soluble pTK protein fragments can also be used in binding assays to detect ligands such as growth and differentiation factors. Once these ligands are identified, they may be altered or modified to inhibit or
25 enhance kinase activity. For example, the ligands may be modified or attached to substances that are toxic to the cell, such a ricin, thus destroying the target cell. The substance may be a super-activating substance which, after binding to the pTK, may substantially increase the kinase
30 activity, or activate other growth factors.

-13-

pTk genes of the present invention would also be useful to develop diagnostic tools for in vitro screening assays for ligands such as growth factors or differentiation factors that inhibit or enhance kinase
5 activity. The proteins encoded by the pTK genes can also be used in such assays, or as immunogens to produce monoclonal or polyclonal antibodies to be used in such assays.

Such antibodies can also be used in methods of
10 treating conditions in which an individual would benefit therapeutically if protein tyrosine kinase activity could be modified, such as increasing platelet production in bleeding disorders.

The present invention will now be illustrated by the
15 following Exemplification, which is not intended to be limiting in any way.

Exemplification: The Identification and Isolation of the pTK Genes

To facilitate the isolation and identification of
20 these novel pTK genes, two sets of DNA probes were used. (See Table).

The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK
2 (SEQ ID NO:2). These sequences were used as polymerase
25 chain reaction (PCR) primers, using standard PCR techniques, to amplify tyrosine kinase DNA segments.

The second set consisted of two oligonucleotide sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4) selected from the highly conserved regions of the
30 catalytic domains of the c-kit subgroup of protein

-14-

tyrosine kinases. These sequences were also used as polymerase chain reaction primers in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK
5 primers were identified, isolated and subsequently sequenced using known laboratory techniques.

TABLEFirst Round of Amplification

PTK1

10 CGGATCCACAGNGACCT

PTK2

GGAATTCCAAAGGACCAGACGTC

Second Round of Amplification

PTK3 (kit family specific)

15 CGGATCCATCCACAGAGATGT

PTKKW (kit family specific)

GGAATTCCTTCAGGAGCCATCCACTT

Equivalents

Those skilled in the art will recognize, or be able
20 to ascertain using no more than routine experimentation,
many equivalents to the specific embodiments of the
invention described herein. Such equivalents are intended
to be encompassed by the following claims.

-15-

CLAIMS

The invention claimed is:

1. Isolated DNA of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA
5 encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.
2. Isolated DNA of Claim 1 having a nucleotide sequence selected from the group of nucleotide sequences
10 consisting of:
 - a) SAL-S1 (SEQ ID NOS:5 and 7);
 - b) SAL-D4 (SEQ ID NO:7);
 - c) LpTK 2 (SEQ ID NOS:9 and 19);
 - d) LpTK 3 (SEQ ID NO:11);
 - 15 e) LpTK 4 (SEQ ID NOS:13 and 21);
 - f) LpTK 13 (SEQ ID NO:15);
 - g) LpTK 25 (SEQ ID NO: 22 and
 - h) HpTK 5 (SEQ ID NO: 23).
3. Isolated DNA of Claim 1 which encodes an amino acid
20 sequence selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NOS:6 and 18);
 - b) SAL-D4 (SEQ ID NO:8);
 - c) LpTK 2 (SEQ ID NOS:10 and 20);
 - d) LpTK 3 (SEQ ID NO:12);
 - 25 e) LpTK 4 (SEQ ID NO:14);
 - f) LpTK 13 (SEQ ID NO:16);
 - g) HpTK 5 (SEQ ID NO:24);

-16-

- h) bpTK 1 (SEQ ID NO:25);
- i) bpTK 2 (SEQ ID NO:26);
- j) bpTK 3 (SEQ ID NO:27);
- k) bpTK 4 (SEQ ID NO:28);
- 5 l) bpTK 5 (SEQ ID NO:29); and
- m) bpTK 7 (SEQ ID NO:30).

4. Isolated DNA of human megakaryocytic origin which comprises a DNA fragment whose sequence encodes an amino acid sequence present in the catalytic domain
10 of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.

5. Isolated DNA of Claim 4 which encodes an amino acid sequence selected from the group consisting of:

- a) SAL-S1 (SEQ ID NO:6);
- 15 b) SAL-D4 (SEQ ID NO:8);
- c) LpTK 2 (SEQ ID NO:10);
- d) LpTK 3 (SEQ ID NO:12);
- e) LpTk 4 (SEQ ID NO:14); and
- f) LpTK 13 (SEQ ID NO:16).
- 20 g) HpTK 5 (SEQ ID NO:24);
- h) bpTK 1 (SEQ ID NO:25);
- i) bpTK 2 (SEQ ID NO:26);
- j) bpTK 3 (SEQ ID NO:27);
- k) bpTK 4 (SEQ ID NO:28);
- 25 l) bpTK 5 (SEQ ID NO:29); and
- m) bpTK 7 (SEQ ID NO:30).

-17-

6. A homogeneous protein of human megakaryocytic origin which includes an amino acid sequence exhibiting sequence homology with the catalytic domain of tyrosine kinases of the c-kit family.
- 5 7. A homogeneous protein of Claim 6 in which the amino acid sequence is selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NO:6);
 - b) SAL-D4 (SEQ ID NO:8);
 - 10 c) LpTK 2 (SEQ ID NO:10);
 - d) LpTK 3 (SEQ ID NO:12);
 - e) LpTk 4 (SEQ ID NO:14); and
 - f) LpTK 13 (SEQ ID NO:16).
 - g) HpTK 5 (SEQ ID NO:24);
 - 15 h) bpTK 1 (SEQ ID NO:25);
 - i) bpTK 2 (SEQ ID NO:26);
 - j) bpTK 3 (SEQ ID NO:27);
 - k) bpTK 4 (SEQ ID NO:28);
 - l) bpTK 5 (SEQ ID NO:29); and
 - 20 m) bpTK 7 (SEQ ID NO:30).
8. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FLT/FLK family of protein tyrosine kinases.
9. A protein of Claim 8 encoded by the nucleotide
25 sequence (SEQ ID NO:5).
10. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO:6).

-18-

11. A protein of Claim 8 encoded by the nucleotide sequence (SEQ ID NO:17).
12. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO: 18).
- 5 13. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FGF receptor family of protein tyrosine kinases.
14. A protein of Claim 10 encoded by the nucleotide sequence (SEQ ID NO:7).
- 10 15. A protein of Claim 10 encoded by the amino acid sequence (SEQ ID NO:8).
16. A protein of human megakaryocytic origin which exhibits significant sequence homology with the NGF receptor family of protein tyrosine kinases.
- 15 17. A protein of Claim 14 encoded by the nucleotide sequence (SEQ ID NO:11).
18. A protein of Claim 14 encoded by the amino acid sequence (SEQ ID NO:12).

-19-

19. A DNA expression vector containing a DNA sequence of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases.
20. The DNA expression vector of Claim 17 containing a DNA sequence selected from the group consisting of:
- a) SAL-S1 (SEQ ID NO:5);
 - b) SAL-D4 (SEQ ID NO:7);
 - c) LpTK 2 (SEQ ID NO:9);
 - d) LpTK 3 (SEQ ID NO:11);
 - e) LpTK 4 (SEQ ID NO:13); and
 - f) LpTK 13 (SEQ ID NO:15).
 - g) LpTK 25 (SEQ ID NO: 22 and
 - h) HpTK 5 (SEQ ID NO: 23).
21. A cell transformed by the expression vector of Claim 17.

1/35

sals1 (160 bases)
FLK1-LIKE

1	PTK1/3 PRIMERS	21	41
5' <u>ggatcctgtgcatcagtgacttagggctaggaacattctgtgtcggaagcgacgtggt</u>			
		D P V H Q A L R A R N I L L S E S D V V	
61		81	101
gaagatctgtgactttggccttggccgggacatctacaagagacccagctacgtccgcaa			
		K I C D F G L A R D I Y K D P S (Y) V R K	
121		141	
gcattgccggctgccctgaagtggatggcgccagattc 3'			
		H A R L P L K W H A P E F	

FIGURE 1

sald4 (147 bases)-
FGFR-LIKE

1	PTKI/3 PRIMERS 21	41
5' <u>ggatccattcacagagac</u> cttagcagcacgcaacatcctgggtctcagaggacctggtaacc		
		G S I H R D L A A R N I L V S E D L V T
61	81	101
aagtcagcgactttggcctggccaaagccgagcggaagggttagactcaagccggctg		
		K V S D F <u>G L A K A E R K G L D S S R L</u>
121	PTKKW PRIMER 141	
<u>ccg</u> tcaaatggatggctcccgaattc 3'		
		P V K W H A P E F

FIGURE 2

3/35

LpTK2

GTTGGAATTCCTTCGGGCGCCATCCATTTACCGGCAGCTTTATTTTCGTGTCTAGATTCA
TAGATGTCTTCATTATCTACCTTAAAACTCTGGCAAGTCCAAATCTGCTACTTTGTAG
ATATTATGTTACCAACGAGGACATTCCT

FIGURE 3A**LpTK3**

GTGCACAGGGATCTCGCGGCTCGGAACATCCTCGTCGGGGAAAACACCCCTCTCGAAAGTT
GGGGACTTCGGGTTAGCCAGGCTTATCAAGGAGGACGTCTACCTCTCCCATGACCACAAT
ATCCCCTACAAATGGATGGCCCCCTGAGGGAA

FIGURE 3B**LpTK4**

GTTACCGGAGATCTCAAGTCCAACAACATTTTGCTGCTGCAGCCCATTGAGAGTGACGAC
ATGGAGCACAAGACCCTGAAGATCACCGACTTTGGCCTGGCCCGAGAGTGGCACAAAACC
ACACAAATGAGTGCCGC

FIGURE 3C**LpTK13**

GTCAATCGTGACCTCGCCGCCCCGAAATGTGTTGCTAGTTACCCAACATTACGCCAAGATC
AGTGATTTCGGACTTTCCAAAGCACTGCGTGCTGATGAAACTACTACAAGGCCAGACC
CATGGAAAAGTGGCCTGTCAAGTGGTACGCTCCGGAATGCATCAACTACTACAAGTTCTCC
AGCAAAAGCGATGTCTGGTCCTTTGGAATTC

FIGURE 3D

5/35

701 TTGGAACGGG GATTCCCGGT GCCAAGAGTG ACGTAAGTAC CGCCTATAGA GGTATAGGC CCALTTGGCT TCCTTAGAAC GCGGCTACAA TTAATACATA
AACCTTGGCG CTAAAGGGCA CGGTTCTCAC TGCATTCTATG GCGGATATCT CAGATATCCG GGTCAACCGA AGCAATCTTG CCGCGATGTT AATTATGTAT

801 ACCTTATGTA TCATACACAT ACGATTTAGG TGACACTATA GANTACATC CACTTTGCTT TTCTCTCCAC AGGTGTCAC TCCCAGGTC AACGCACTT
TGGATATCAT ACTATGTGTA TGCTAATCC ACTGTGATAT CTTATTCTAG GTGAAACGGA AAGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGGA

901 CGGTTCTATC GATTGAATTC CCGGGGATC CTCAGAGAT CCGTGACCT CGAGATCCAT TGTGCTGGCG CGGATTCCTT ATCACTGATA AGTTGGTGGA
GCCAAGATAG CTAACCTAAG GGGCCCCCTAG GAGATCTTA GGAAGCTGGA GCTCTAGGTA ACACGACCGC GCTTAGAA TACTGACTAT TCANCCACCT

1001 CATATTATGT TTATCAGTGA TAAAGTGTA ACGATGACAA AGTGCAGCG GAATACAGTG ATCCGTGGCG CCGTAGACCT GTTGACGAG GTCCGGCTAG
CTATATATCA AATAGTCACT AATTCACTT TCCTACTGTT TCACCGTGG CTTATCTCAC TAGGCACGGC GGGATCTGGA CACTTGCTC CAGCCGCTC

1101 ACGGTCTGAC GACAGGCNA CTGGCGGAAC GGTGGGGGT TCAGCAGCGG GCGTTTACT GGCATTTCAG GAACAGCGG GCGCTGCTCG ACGCACTGGC
TGCCAGACTG CTGTGGGTTT GACCGCCTTG CCACCCCGCA AGTCGTGGCG CCGGAANTGA CCGTGAAGTC CTTGTTGCG CCGGACGAGC TGGGTGACCG

1201 CGAAGCCATG CTGGGGAGA ATCATAGCAC TTGGTGCGG AGGCCGACG ACGACTGGCG CTCATTCTTG ACTGGGANTG CCGCGAGCTT CAGGCAGGGG
GCTTCGGTAC GACCGCCTCT TAGTATCTGT AGGCCACCG TCTCGGCTGC TCGTGCAGCG GAGTAAGAC TGACCTTTAC GGGCGTGGAA GTCCGTCCCG

1301 CTGCTCGCT ACCGCCAGCA CAATGGATCT CGAGGATCT TCCATACCTA CCAATTCTGC GCTGTCAGGT CCGGGCCCA CTACTCTTG ATGTATTACT
GACGAGCGGA TGGCGGTGCT GTTACCTAGA GCTCCCTAGA AGGTATGNT GGTCAAGACG CGACGTCCA GCGCGGCGT GATGAGAAAC TACATANTGA

FIGURE 4B

1401 CATATTACCA AGGAATAACT GGCGGGCACA GGTCAGGTG CTGAAGGGAC ATTGTGAGAA GTGACCTAGA AGGCAAGAGG TGAGCCCTCT CTCACGCTGG
 GTATATGGT TCCATTATGA CCGCGGTGT CCGAGTCCAC GACTTCCCTG TAAACACTCT CACTGGATCT TCCGTCTCC ACTCGGAGGA ACTCGCGACC

1501 CATAGGGGC GGTGAGGGC TCTTGGTCA AGCAGTAACG CCACTGTCTG GGAAGGCACC TGTACTCAG CAGACCATGA AAGGGCTCT CCTTTTCCT
 GTATCCCG CGAATCCCG AGAATCCAGT TCGTCAATCC GGTCAACAGC CCTTCCCTGG ACAATGATC GTCTGTACT TCCCGCAGA GGGAAAGGAA

1601 GCAGCACTCA GGGAACTCT TCTCCACCA GCTTCTGTG GGAGGCTGGA TATTATCAG GCTGCGCGG AGTCATCCG AGCCTAACG CCTCCCTCTG
 CCTCCTCAGT CCTTGTGAG ACGAGCTGGT CGAAGAACAC CCTCCGACCT ATANTAGTC CGGACGGGG TCAATAGGCC TCGGATTCG GCAGGGACAC

1701 GGTCTCAGT GGTACACTC CTGTCCACT TCTATGCTC TCTTGGCTC CTTGTCTCTG TCTGAGATA GGAGAAGAA TAGGGAAGT
 CACGAGTCA CCAGTGTGAG GAACAGTGA AATACGAGG AGAACGGAG ACCCTTCAA CATCATCTAT CTTCTCTT ATCGCTTCA

1801 CTTAAGTCT TTGATCTTC TTATAGTGC AGAGAAGAA TCTGACGTA TCTGCTTC TCTCTCTG CTTAGCTAG CTGAAGCGC TTTCTGTCT
 GAATTCAGA AACTAGAAAG AATATTCAG TCTCTCTT ACCACTGCAT ACCAGGGAG AGAGAGAGC GAATCGATG O R P G S E Q R
 349

1901 ATACCTGCTC TCTATCTGT CACTCTCTC CGAGGCCAGC ACCATCCAC TGTCTGTCTG GTTGTCCACA GAGCCTTGT AGTCTGTGG GGTCTGGG
 TATGACGAG AGATAGACA GTCTGAGGAG GCTCCGCTG TGTAGGTG ACAGACAGC CAACAGTCT CTCGGAACA TCCAGCAAC CCAGTACCC
 341 Y R S E I Q E C E E S A L V M C S D T Q N D V S Q K Z T T P Z H P

2001 ATTCTCTCA ATGTCTTCT CTTGAGGAA CCAAGGTCT CAGCCCTCT GGCAGGCAC CCGGAAGG ACACCAAT GTATPACTG GCGCCAGGC
 TTAGGAGCT TACAGAGTA GGACTCTCT GTGCCCCA GTGCGGAGA CCGTCCCTG GGCCTTCC TGTGGTCAA CATATGGAC CGCGGTCCG
 308 P E E F T K M R S S G R T E A C R A L C G P F S V W N Y Y R A A L

FIGURE 4C

2101 TGTGGCGCTG CAGGCTTGGC GGGTGTCTCT CAGCGTCAGC CTGGGCGCA TGTGGACAC CTGGGAGAG CTGCCCTCTT CTGAGCTCTG
 ACACCGGAC GTCCGACCG CCGGACAGG GTCCGAGTGC GACCGGTAC ACATCCGGT ACCACTGTG GACGCTCTC GACGGAGAA GACTCGAGAC
 275 S H R Q L S P P S D E A D A Q A I H L A M T S V Q S F S G E E S S Q

2201 AGAGTGGGC GGGGCGATGC AGACTCTCTC TTCTCTTGC AGCCCTTGC CTGGAGCAG GTCCCCGAG ATCTCCACCA GCTCCGAGAA TGCAGCTCTC
 TCTCGACCG CCGCGTACG TCTGGAGAG AGGAGAACG TCCGGGACG GGACTCTGC CAGGGGTCC TAGAGGTGT CGAGGCTCTT ACCTCCAGAG
 241 S S R P A M C V E E E Q L G R G O L L D G L I E V L E S F A P R

2301 GCCTTGGGT CTCCGGACCA GCACTCAGC ATGATGCGG GATGGCGG AGTGGCCAGC TCCGGGCCC TCATCTTGT GCCGTCTCTC AGCCGCTGCG
 CGGAACCCA GAGGCTGTG TACTAGTCCG TACTAGTCCG CATACGCGC TCACCGGTG AGGCCCGG AGTAGGAACA CGGCAGAGAG TCGGCGACCG
 208 A K P D G S W C N L M I R R I A P T A L E P A R M R T G D R L R Q

2401 AGACTCTCTC ATGATCTGC ACCCAGGCT ACGGGAGGC CCCAGAGAG AGATCTCC AGAGAGCAG CCCAAGGAC CACAGCTCAC TCTGGTGGT
 TCTGAGGAG TAATCTACG TGGGTCCCA TCCCTCCG Y P S A G L S F I E W L L V G P S W V D S Q T T
 175 C F E E N I Q V G P Y P S A G L S F I E W L L V G P S W V D S Q T T

2501 GTACACUTTG TCGAAGATGC TTTCAGGCG CATCCACTTC AGGGCAGCC GGGCACTGCC CTGGCGGAG TAGTCGGGT CTTCGAGAT GTCCCGGCA
 CATGTGAGAC AGCTTCTACG AAAGTCCCG GTAGGTGAG TCCCGTCCG CCCGTGACG GAACGCTGC ATCAGCCCCA D K Y I D R A
 141 Y V K D F I S E P A M W K L P L M A S G K R V Y D P

2601 AGGCCAAGT CACAGATCTT CACCACGTCG CTTTCCGACA GCAGATGTT CCAGAGAGC AGGCTCTGT GGTGCACTT TCGGAGCC AGGAATCCA
 TCGGGTTTCA GTGTCTAGAA GTGGTGCAGC GAAGGCTGT CGCTTACA GGCTCTCGG TCCAGAGACA CCTACGTGA AGCCCTCTGG TCCTGAGGT
 108 L G F D C I K V V D S E S L L I N R A A L D R H I C K R S A L P E

2701 TCCCTCTGCG CACCTGGAG CTGTAGCAGA CAAGATCTC CATGTCTCAG GGGCTCAGC ACAGGTCTCT AGCTCTTGG TCTGGAGAG CCGGCTCTGC
 AGGGAGACCG GTGGACCTT GACATCGTCT GTCTAGAG GTACCACTCG CCCGAGTCCG TGTCCAGAG TCGAAGAAC AGACTCTCTC GGGCGAGCG
 75 M G R A V Q F S Y C V L D E M T L P S L W L D E A E Q D P S A R R A

FIGURE 4D

FIGURE 4E

2801 TCCGCCCTCG GTCTTCUNGA ACCGUGGAA GAGGACCTG TCGTGTCTC CCGCCGCTT CCGATCCAG CTGGGAGCT CCACATGGC GCGGAAGCT
 AGCGGGAGC CAGAAGCTT TGGGCGCTT CTCTGGAC AGCAGCAGG GCGCGCGA GCTAGTCTG GACCGCTGA GGTGTACCG CGCTTCGA
 41 C G E T K S P R A P L V R D S S G P R R R A L E V M A R P R
 2901 CCGCGTGTCT CCGGAGACTT CTCTGCGGA TGCACGAGC TGGTGTAGG GCGGAGTC GTCCGCGCA GAGCGGCTC CATTCGCCG CCGCCGCGG
 GCGCGACGA GCGCTCTGA GAGGACGCTT AGGTGTCTG ACCGAGCTC CCGCGCTAG CAGCGGCTT CTCCGCGG GTAAGGGGC GCGCGCGC
 8 G R Q E P S K E
 3001 CCGCGGCGC GCGGCGGCTT CCGCGGCGC GCGGCGGCTT CCGGCGGCTT CCGGCGGCTT GCGGCGGCTT GCGGCGGCTT GCGGCGGCTT
 GCGGCGGCTT GCGGCGGCTT GCGGCGGCTT GCGGCGGCTT GCGGCGGCTT GCGGCGGCTT GCGGCGGCTT GCGGCGGCTT GCGGCGGCTT
 3101 CTTATATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAA GCATTTTTT CACTGCATC TAGTGTGTG TTGTCCAAAG TCATCAATGT
 GAATATACC AATGTTTAT TCGTTATCGT AGTGTATAA GTGTTATTT GTGAAATAA GTGAGGTAG ATCAACACCA ACAGGTTTG AGTAGTACA
 3201 ATCTTATCAT GTCTGGATCG ATCGGGATTT AATCGGGCTT AATCGGGCTT AATCGGGCTT AATCGGGCTT AATCGGGCTT AATCGGGCTT
 TAGATAGTA CAGACCTAGC TAGCCCTTAA TTAAGCGCGT TCGTGTAGC GACTTTTAT GACACTTTC TCTTGAACC AATCCATGGA AGACTCCGC
 3301 AAAGAACCA GTGTGGATG TGTGTAGTT AGGTGTGGA AATGTCCAGC GTTCCGAGC GCTCCGAGC AGCAGAGAT ATGCAAGCA TGCATCTCA TTAGTCAGCA
 TTCTTGTCT CACACCTTAC ACACAGTCA TCCACACCTT TCCAGGCTC GAGGCGCTT TCCGCTTCA TACGTTTCT ACCTAGGTT AATCAGTCT
 3401 ACCAGGTGT GAAATCCCC AGCTTCCCC GAGGCGGAA GTATGCAAG GTATGCAAG GTATGCAAG GTATGCAAG GTATGCAAG GTATGCAAG
 TGTCCACAC CTTTACGGG TCGAGGGGT CGTCCGCTT CATACGCTT CATACGCTT CATACGCTT CATACGCTT CATACGCTT CATACGCTT CATACGCTT

9/35

FIGURE 4F

3501 TCCGCGCCCT ANCTCCGCC AGTTCGCCCC ATTCTCCGCC CCATGGCTGA CTAAATTTT TTAATTATGC AGAGGCCGAG GCGGCTCGG CCTGTAGCT
 AGGCGGGGA TTGAGGGGG TCAAGGGGG TAAAGGGGG GTACCGACT GATTAANAA ATAAATACG TCTCGGCTC GCGCGAGCC GGAGACTCGA

3601 ATTCCAGAG TAGTGAGGAG GCTTTTGG AGGCTAGGC TTTTGCNAA AGCTGTAAAC AGCTTGGCAC TGGCGTGGT TTTACAACTG GTTGACTGGG
 TAAGGTCTTC ATCAGTCCTC CGAANANACC TCCGNTCC AAACGTTTT TCGACAATTG TCGNACCGTG ACCGACCA AATGTGGCA GCACTGACCC

3701 AAAAACTGG GGTATCCCA CTTAATGCC TTGACGACA TCCCCCTTC GCCAGCTGGC GTAAATAGCA AGAGGCCCG ACCGATCGC CTTCCTCAACA
 TTTGGUACC GCANTGGGT GANTTAGCG AACGTCTGT AGGGGGGAG CGTCCGACC CATATCGCT TCTCGGGCG TGGTAGCGG GAAGGTTGT

3801 GTTGGTAGC CTGAATGGG ANTGGGCTT GATGGGTAT TTCTCTTA CGCATCTGG CGTATTTCA CACGCTAC GTCAAGCA CCATAGTACG
 CACGCATCG GACTTACCG GACTTACCG TTACCGGA CTACGCCATA AAGAGGAAT CGGTAGCAC GGTATAGT GGTATCATGC

3901 CGCCCTGTAG CGGCGCATTA AGCGGGCGG GTGTGGTGT TACGGGAGC GTGACGGTA CACTGGCTG GTGAAGGCTC GCGGATCG GCGGAGAA ACCGAAAGAA
 GCGGGACATC GCGCGTAA TCGCGCCGC CACACGACCA ATGCGCTCG CACTGGCTG GCGGATCG GCGGAGAA ACCGAAAGAA

4001 CCTTCTTT CTGCGCACGT TCGCGGCTT TCCCGTCA GCTTAATC GGGGCTCC TTTAGGCTC GATTATAG CTATACGCA CCTGACCCC
 CGAAGGAA GAGCGGTCA AGCGCGGA AGGGCAGT CGAGATTAG CCCCCAGG AATCCCAAG GCTAATCAC GAATGCCGT GAGCTGGG

4101 AAAAACTTG ATTGGGTGA TGGTTCAGT AGTGGGCCAT CGCCTGATA GAGGTTTTT CGCCTTTGA CGTTGAGTC CACTTCTTT AATAGTGAC
 TTTTTCATC TAAACCACT ACCAAGTCA TCACCGGTA GCGGAGCTAT CTGCAAAAA GCGGAAACT GCAACTCAG GTCAAGAAA TTATCACCTG

10/35

FIGURE 4G

4201 TCTTGTCCTCA AACTGGAACA ACACTCAACC CTATCTCGGG CTATCTCTTT GATTATTAAG GGATTTGCC GATTTCGGCC TATTTGGTAA AAATGAGCT
 AGAACAAAGT TTGACCTTGT TGTGAGTGG GATAGAGCCC GATANGAANA CTAATATATC CTAATAACGG CTAAAGCCGG ATACCAANTT TTTTACTCGA

4301 GATTTAACAA AAATTAAACG CGAATTTTAA CAAATATTA ACCTTACAA TTTATGGTG CACTCTCAGT ACAAATCTGCT CTGATGCCCG ATACTTAAGC
 CTAAATTTGT TTAAATTCG GCTAAATTT GTTTATANT TGCANATGTT AATATACCAC GTGAGAGTCA TGTATAGCGA GACTACGGCG TATCAATTCG

4401 CAACCTCCGT ATCGCTACGT GACTGGGTCA TGGCTGGCC CGACACCCG CCAACACCCG CTGACGGCGC CTGACGGGCT TGTCTGCTCC CGGCATCCGC
 GTTGAGGCCA TAGCGATCCA CTGACCCAGT ACCGACGGCG GGCTGTGGC GGTGTGGCG GACTGCCCG GACTGCCCG ACAGACGAGG GCCGTAGCGC

4501 TTACAGACAA GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGGTFTT CACCGTCAATC ACCGAAACGC GCGAGGCAGT ATCTTGAG ACCAAAGGCC
 AATGTCGT CGACACTGGC AGAGGCCCTC GAGTACACA GTCTCCAAA GTGGCAGTAG TGGCTTGG CGCTCGTCA TAAGAACTTC TGCTTTCCCG

4601 CTCGTGATAG GCTTATTTT ATAGGTAAAT GTCATGATAA TAATGCTTTC TTAGACGTCA GGTGGCAGT CCACCGTGA AGCCCCCTT ACACGGGCT TGGGGATAA
 GAGCACTATG CGGATAAATA TATCCAAATA CAGTACTATT ATTACCAAG AATCTGCAGT CCACCGTGA AGCCCCCTT ACACGGGCT TGGGGATAA

4701 GTTATTTT CTAAATACAT TCAATATGT ATCCGCTCAT GAGCAATAA CCTGTATAA TGCCTCAATA ATATTGAAA AGGAAGATTA TGATATTCA
 CAAATAAANA GATTATGTA AGTTATACA TAGGCGAGTA CTCTGTAT GGGACTATT ACGAAGTAT TATAACTTTT TCCTCTCAT ACTCATAGT

4801 ACATTTCCGT GTCGCCCTTA TTCCCTTTT TCGGCGATTT TGCTTCTCTG TTTTGTCTCA CCCAGAAACG CTGGTGAAAG TAAAGATCC TGAAGATCAG
 TGTAAAGGCA CAGCGGGAT AAGGGAAAA ACGCCATAA ACGBAAGGAC AAAAACGAGT GGTCTTTTC GACCACCTTC ATTTCTACG ACTTCTAGTC

FIGURE 4H

4901 TTGGGAGCAG GAGTGGGTTA CATCGAACTG GATCTCAACA GCGTAAGAT CCTGAGAGT TTTCGCCCG TTTCGCCCG TCCATGATG AGCATTCTT
AACCCACGTG CTCACCCCAAT GTAGCTTGAC CTAGAGTTGT CGCATCTTA GGAATCTCA AAGCGGGGC TTCTGCANA AGTTACTAC TCGTGAANA

5001 AAGTCTGCT ATGTGGCGG GTATTATCC GTGATGACG CAGGCAAGAG CAATCTGGTC GCGCATACA GCGCATACA CTATCTCAG ATGACTTGG TTGACTACTC
TTCAAGACGA TACACCGGC CATATAGGG CACTACTGG GCGCTCTC GGTGAGCCAG CCGGTATGT GATNAGATC TTACTGAAAC AACTCATGAG

5101 ACCACTCACA GAAAGCATC TTACGGATG CATGACACTA AGGAAATAT GCAGTGTGG CATTAACCATG AGTGAZACA CTGGGCCAA CTACTTCTG
TGGTCACTGT CTTTCGTAG AATGCCATC GTACTGTCT TCTTAAAT CGTACGACG GTATTGGTAC TCACTATGT GACGCGGT GAAATGAAG

5201 ACAACGATCG GAGGACCGAA GGAGCTAACC GCTTTTTCG ACAACATGG GATCATGTA ACTGCTTGG ATCGTTGGA ACCTGAGCTG AATGAAGCA
TCTTGCTAGC CTCCTGCTT CTTGATGG CGAANAACG GCTTTTTCG TGTGTACCC CCTAGTACAT TGAGCGGAAC TNGCAACCT TGGCTCGAC TTACTTCTG

5301 TACCAACGA CGAGCGTAC ACCACGATG CAGCAGCAAT GGCAACAAG TTGGCAAC TATTACTGG CGAATCTT ACTTAGCTT CCGGCCAACA
ATGGTTTGT CCTCGCACTG TGGTGTACG GTCTGTCTTA CCGTTGTTC AACGCTTTC ATATTGACC CTTGATGNA TGAGATGNA GGGCGTGT

5401 ATTAATAGAC TGGATGAGG CGGATTAAGT TGCAGGACCA CTCTGCGGT CCGCTTCC GGTGGCTGG TTATTCCTG ATAACTGG AGCGGTGAG
TAATATCTG ACCTACCTCC CCTATTTCA ACCTCTTGT GAGACCGCA GCGGGGAGG CCGACCGACC AATNACGAC TATTAGAC TCGGCCACTC

5501 CTTGGCTCTC GCGGTATCAT TGCAGCACTG GGGCAGATG GTAGCCCTC CCGTATCGTA GTTATCTACA CCACGGGGAG TCAGGCACT ATGATGAC
GACCCAGAG CGCCATAGTA ACCTCGTAC CCGGTCTAC CATTCGGAG GGCATAGCAT CAATAGATGT CTTGCCCCC ACTCGGTGA TACTACTTG

FIGURE 4I

5601 GAAATAGACA GATCGCTGAG ATAGGTGCGT CACTGATTAA GCATTGGTAA CTGTGAGACC AAGTTTACTC ATATATACCT TAGATTGATT TAAACTTT
CTTTATCTGT CTACCGACTC TATCCACGGA GTGACTAATT CGTAACCATY GACAGTCTGG TTCAATGAG TATATATGAA ATCTAACTAA ATTTTGAAT

5701 TTTTATATTT AARAGATCT AGGTGAAGAT CCTTTTIGAT AATCTATGA CCAAAATCC TTAACGTGAG TTTTCCTTCC ACTGAGCCTC AGACCCCGTA
AAAAATTAAA TTTTCTCTAGA TCCACTICTA GGAANAATA TAGAGTACT GTTTTAGG AATTCCTACT ANAGCAAGG TCACTCGCAG TCTGGGGCAT

5801 GAAAAAATCA AUGATCTTC TTGAGATCTT TTTTCTTGC GCGTAATCTG CTGCTTGCAA ACAAAAAAC CACCGCTACC ACCGTGCTT TGTTCGCGG
CTTTCTAGT TTCTTAGAG AACTCTAGGA AAAAAAGCG CGCATTAGAC GACGAACGTT TGTTTTITG GTGCCGATGG TCGCCACCAA ACRAACGEC

5901 ATCAAGAGCT ACCACTCTT TTTCGAGG TAACTGGCT CAGCAGAGCG CAGATACCAA ATACTCTCTT TCTAGTGTAG CCGTAGTTAG GCCACCACTT
TAGTCTCGA TGGTTGAGAA AAAGGCTTCC ATTGACCGAA GTCTCTTCCG CTCATGCTT TATGACAGGA AGTACATC GGCATCAATC CCGTGTGAA

6001 CAAGAACTCT GTAGCACCGC CTACATACCT CGCTCTGCTA ATCTGTATC CAGTGGCTGC TGGCAGTGC GATAGTCTT GTCTTACCAG GTTGGACTCA
GTCTTTGAGA CATCGTGGC GATGTATGGA GCGAGACGAT TAGGACAATG GTACCGACG ACCTTACCG CTATTCAGCA CAGATGGCC CAACCTGAGT

6101 AGACGATAGT TACCGGATTA GCCCGACCG TCGGGCTGAA CCGGGGCTC GTGCACACAG CCCAGCTTG ACCGAACGAC CTACACCGAA CTGAGATACC
TGTGCTATCA ATGGCTATT CGCGTGGC AGCCCGACTT GCCCCCAAG CAGCTGTCT CAGCTTCTG GATGTGGCTT GACTCTATGG

6201 TACAUCGTGA GCATTGAGAA AGCGCCACGC TTCCCGAAGG GAGNAAGCG GACAGGTATC CGGTAAAGCG CAGGCTCGA ACAGGAGAGC GCACGAGCG
ATGTCGCACI CGTAACCTT TCGCGGTGG AAGGCTTCC CTCTTCCG CTGTCCATAG GCCATTCGCC GTCCAGCTT TGTCTCTCG CGTCTCTCC

FIGURE 4J

6301 GCTTCCAGGG GGAACCCCT GGTATCTTA TAGTCTGTC GGGTTGCGC ACCTCTGACT TGAGGTCGA TTTTGTGAT GCTCTCAGG GCGGCGGAGC
CGAAGGTCCC CTTTGCGBA CCATAGAAAT ATCAGGACAG CCGAAGCGG TGGAGACTGA ACTGCGAGCT AAAACACTA CGAGCAGTCC CCGCGCCTCG

6401 CTATGGAAA ACGCCAGCA CCGGCGCTTT TTACGGTTC TGGCGTTTG CTGGCCTTTT GCTCAGATG TCTTCTCTCC GTTATCCCT GATTCTGTGG
GATACCTTTT TCGGTCTGTT GCGCGGAAA ANTGCCAAG ACCGGAAC GACCGMAA CTGGCTGCTC CGAGTGATCA AGAAGGAGC CAATAGGGA CTAGACACC

6501 ATACCGTAT TACCGCCTT GAGTGAGCTG ATACCGCTCG CCGAGCCGA ACCACCGAGC GCAGCGATC AGTACGGAG GAACCGGAG ACGCCCAAT
TATTGGCATA ATGGCGAAA CTCACTCGAC TATGGCGAGC GCGCTGGCT TCTTGGCTCG TCACTGCTC CTTCGCTTC TCGCGGTTA

6601 ACGCAACCG CCTCTCCCG CCGGTGGUC GATTCATTA TCCAGCTGC ACGACAGTT TCCGACTGG AAGCGGCA GTGAGCGCA CGCAATTAT
TCCGTTTGGC GGACAGGGC GGCNACCG CTAAGTAAT AGTGGAGCG TGTGTCCA AGGCTGACC TTTCGCGCT CACTCGCTT CGGTAAATTA

6701 GTGAGTACC TCACTCATTA GGCACCCCA GCTTACACT TTATGCTTC GCTCCTATG TTGTGGAA TTGTGAGCG ATACATTT CACACAGGA
CACTCANTGG AGTGAATAT CCGTGGGTC CGAATGTGA AATACGAGG CCGAGCATAC AACACCTT AACACTGCC TATGTAAA GTGTGCTT

6801 ACAGCTATGA CCATGATTAC GAATTAA
TGTGATACT GGTACTAATG CTTAAT

FIGURE 5A

1 TTCCAGCTCG CCGACATTCG ATTATTTGACT AGTATATTTAAAT AGTATATTTAAAT TACGGGCTCA TTAGTTTCATA GCCCATATAT GCCCATATAT GGAGTTCCGC GTTACATATAC
 AACCTCGAGC GGGCTGTAAAC TAATTAACCTGA TCAATTAATTTA TCAATTAATTTA ATGCCCCCAAT AAATCAAGTAT CCGGTATATA CCTCAAGGCG CAATGTATTTG

101 TTACGGTAA TGGCCCGCCT GGCTGACCGC CCACGACCC CCGCCCATTC ACCTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTTCCA
 AATGCCATTT ACCGGGCGGA CCGACTGGCG GGTTCCTGGG GCGGGTAAAC TGCAGTTATT CTGCATACA AGGTATCAT TCGGTATATC CCTGAAAGGT

201 TTGACCTCA TGGGTGGAGT ATTTACGGTA AACTGCCAC TTBGCAGTAC ATCAAGTGA TCATATGCC AGTACGCCCC CTATTGACGT CAATGACGGT
 AACTGCAGTT ACCCAGCTCA TAAATGCCAT TTGACGGGTG ACCGTTCATG TAGTTACAT AGTATACGGT TCATGCGGGG GATAACTGCA GTTACTGCCA

301 AAATGGGCG CCTGGCATTA TGCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTACG TATTAGTCAAT CGCTATTACC ATGTTGATGC
 TTTACCGGCG GGACCGTAAT ACGGTCAATG TACTGGATAT CCTGAAAGG ATGAAACGTC ATGTAGATGC ATAATCAGTA GCGATAATGG TACCACCTACG

401 GGTTTTGGCA GTACATCAAT GGGGTGGAT AGCGTTTGA CTCACGGGA TTTCCAAGTC TCCACCCCAT TGACGTCAAT GCGAGTTTGT TTGGCACCA
 CCNAAACCGT CATGTAGTTA CCGGCACCTA TCGCCAAACT GAGTCCCCCT AAAGTTTCAG AGGTGGGATA ACTGCAGTTA CCTCAAAACA AAACCGTGGT

501 AAATCAACGG GACTTTCCAA AATGTGCTAA CAACTCCGC CCATTGACGC AAATGGGCG TAGGCTGTA CCGTGGGAGG TCTATATAAG CAGAGCTGCT
 TTTAGTTGCC CTGAAGGTT TTACAGCAAT GTTGAGGCG GGTAACTGC TTTACCCGCC ATCCGCACAT GCCACCCCTCC AGATATATTC GTCTCGAGCA

601 TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT CCATAGAAGA CACCGGAGC GATCCAGCCT CCGCGGCCCG GAACGGTGCA
 AATCAGTTGG CAGTCTAGCG GACCTCTGCG GTAGGTGCGA CAAAACTGGA GGTATCTTCT GTGCCCCTGG CTAGGTGCGA GCGCGGCCGC CTTGCCACGT

701 TTGGAAACCG GATTCGCCGT GCCAAGAGTG ACCTAAGTAC CGCCTATAGA GTCTATAGGC CCACTTGGCT TCGTTAGAAC CCGGCTACAA TTAATACATA
 AACCTTGGCG CTAAAGGCGA CCGTTCTCAC TGCATTCATG CCGGTATCT CAGATATCCG GGTGAACCGA AGCAATCTTG CGCCGATGTT AATTATGTAT

FIGURE 5B

801 ACCTTATGTA TCATACACAT ACGATTTAGG TGACACTATA GATACACATC CACTTTGCCT TTCTCTCCAC AGGTGTCCAC TCCAGGTCC AACTGCACCT
 TGGAAATACAT AGTATGTGTA TGCTAAATCC ACTGTGATAT CTTATTGTAG GTGAAACGGA AAGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGGA

901 CGGTTCTATC GATTGAATTC CCGGGGATC CTCTAGAGAT CCCTCGACCT CGAGTCGACT TTTTGTAGG CCAAGGGTA CTCTTTTTC
 GCCAAGATAG CTNACTTAAG GGGCCCTAG GAGATCTTA GGGAGCTGA GCTCAGCTGA AAAAAA AAAACATCC GTTTCCCAT GAAGAAAAAG

1001 TTTATTAAAT ACTCAGAAGT CTAGGCCACA GCAATCTACT GTTCTCTCT CATTTTCTA AACTATTTTG ATACCTATTT CTCAGACTTT ATGGGCTATT
 AAATAATTA TGAGTCTTCA GATCCGGTGT CATTAGATGA CAAGAGGAGA GTAAAGGAT TTGATNAAC TATGGATAA GACTCTGAA TACCCGATNA

1101 AGACATTTCT CACATTTCCA TAGATAATA CTGATCCGTT TTGCAACCTG ATTCCTAATA TTAAGAGATT AAACTAATG TATATGACTC TCAGTTGACA
 TCTGTAAAGA GTGTAAAGGT ATCTATTATT GAGTAGGCA ACGTTGGAC TAAGAGTTAT ATTCTCTAA TTTGATTAC ATATACTGAG AGTCMACTGT

1201 CATACTGAAG TACAGAAAA TTCCATCAAT TCTCTCTGCA AAATGAAAA GACTTCGTT TCTCAACAGC TGCATCATTT TTTATGCTAT AGAAAAAT
 GTATGACTTC ATGTCTTTT AAGGTAGTAA AGGAGACGT TTTACTTTTT CTGAAGCANA AGAGTTGTG ACGTAGTAAA AAAATACGTA TCITTTTTTA

1301 GTGCAATTAC TCCAAGTACA ATCAAGTCAT TTAACATGGC TTTACCATCA TTGTAGTTAC AGGATATTTT AAAAGAGAAA AAAAATCTC AAAGCACAGG
 CACGTTAATG AGGTTCATGT TAGTTTCAGTA AATTGTACCG AANTGGTAGT AACATCAATG TCCATATAAA TTTCTCTTT TTTTGTAGG TTTCGTGTCC

1401 TCTGTCTGTG CAGCAAGCA ATCAATTTCC TTCATAATTA CAGCCTGATG GATTCAGCA ATCTGAGGA TAAATGAATA CCACTCTAAT CAGTAAACAG
 AGGACGACAC GTCGTTTCGT TAGTTTAAAGG AAGTATTATT GTGGGACTAC CCTAAGTCGT TAGACTCCTT ATTACTTATT GGTGAGATTA GTCATTGTG

1501 GAATATGCTA CAACAGTCAC TGAGTAAAA TTGACATATC ATCTGTGAT TCTCTTGATC GACATTTCAA ACATAAATG GAATGTAAAG TATCTCTTNA
 CTTTTACGAT GTTGTCAAGT ACTCATTTTT AACCTGATG TAGACAATA AGAGAACTAG CTGTAAAGTT TGTATTATAC CTTTACATTC ATAGAGATTT

FIGURE 5C

1601 AAGAGAAAT AACTTGGTTT AGTGTGCTTA ATTTTACCAG GCAGTGAAGG AATTATATAT CACCTTGACT GTCTGCACT GTTGCCCACT CAATAAATG
 TTCTTTTTA TTGAACCAAA TCACACGAAT TAAATGGTC CGTCACTCCT TTAATATATA GTGGAACCTGA CAGGACCTCA CAACGGGTCA GTTATTTTAC

1701 CACAAATAT CTTTTTCATA ATACATGGCC AACTTTATCC TATCATTGA ATATGTCAGG ATAACTGAT TGTGCAGTTG GTTGATAACA TTGTATTTTG
 GTGTTTATTA GAAAAAGTAT TATGTACCGG TTGAAATAGG ATAGTGAAT TATACAGTCC TATTTGACTA ACACGTCAAC CAATATTGT AACATAAAC

1801 GAATGGATTA TTGGAATTTG TTTTGCTACT TTATTTATTTG ATATCTTCT CCAGTGTTCA TCTTATGAAG TTATTTGCAT CTGAATATGA AGATCTGTT
 CTTACCTAAT AAACCTTAAC AAACGATGA AAATATAAAC TATAAGAGA GGTCACAAAT O R I F N N A D S Y S S D T
 506

1901 TCAAAATAGT CTTCAGTTT CCAACGCACT GTCTCAAAATG TAGGTCTGTC CTTAGGCTCT GCATTCAGC ACTCCAACAT GATGTTGTA AATGCTGTG
 AGTTTATCA GNAGTTCAA GGTTCGCTCA CAGAGTTTAC ATCCAGCAAG GAATCCGAGA CGTAAGGTG TGAGGTGTA CTACAACAT TTAAACGACAC
 492 E F Y D E L K W R L T E F T P R E K P E A N W C E L M I N Y F Q Q

2001 CACAGTTGGA TGGTTGCGA AGTCTATAGT TTTGAGCCAA CATCTGGAT ACCTGGGCAC CTGTCAFACC ACTGTAAGGC ATTTGCCAT AAGTAATGAT
 CTGTCNACCT ACCNACGCT TCAGATATCA AAACCTCGTT GTAGACCTAA TGGACCCGTT GACAGTATGG TGACATTCG TAAACGGTA TTCATTTACTA
 459 P C N S P Q P L R Y N Q A L M Q I V Q A G T M G S Y P M K G Y T I I

2101 TTCATAAGA AGGATTCCAA ATGACCATAC ATCGGACTTA ATGCTGAAT TATTAATAGC AATGGCTTC GGGCAGTCC ACTTCACCG CAGCTTTAT
 NAGTATTTCT TCCTAAGGTT TACTGATATG TAGCCTGAAT TACGACTTAA ATAATGATGC TTACCGAAGC CCGGTGTCAGG TGAAGTGCC GTCGAATTA
 425 E Y L L I G P S W V D S K I S P K N S R I A E P A T W K V P L K I

2201 TCGTGTCTAG ATTCATAGAT GTCTTCAATTA TCTACCTTAA AACTCTGGC AAGTCCAAA TCTGCTACTT TGTAGATATT ATGTTACCA ACAGGACAT
 AGCACAGATC TAAGTATCTA CAGAGTATAT AGATGGAAT TTGAGACCG TTCAGGTTT AGACGATGAA ACATCTATAA TACAAGTGT TGTCTCTGTA
 392 E H R S E Y I D E N D V K F V R A L G F D A V K Y I N H E G V L V

2301 TTCTGGCAGC CAGATCTCTG TGAATGTAGT TCCGAGACTC CAGATAGGCC ATTCCAGAG CAACCTGTGC CGCCATGTCT ACCTGTTGAG TCAGATGGAT
 AAGACCGTCG GTCTAGAGAC ACTTACATCA AGGCTCTGAG GTCTATCCG TTAGGTCTGC GTTGACACG GCGGTACAGA TGGACAATC AGTCTACCTA
 359 N R A A L D R H I Y N R S E L Y A M Q S A V Q A A M D V Q Q T L H I

FIGURE 5D

2401 TTTTGATCCA GTGTCATTTT GGAGATATTC TTGCAGACTT CCAATGCTCA TCAACTCTGT ANTAATATAA ATTGGATCTT CTAAAGTGCA AACAGCATAA
 AAAAACTAGGT CACAGTAATA CCTCTATTAAG AACGTCTGAA GGTAACAGAT AGTTGAGACA TTATTATATT TAACCTAGAA GATTCACGT TTGTCGTATT
 325 K S G T D N Q L Y E Q L S G H R M L E T I I Y I P D E L T C V A Y

2501 AGCTGGATAA GCTTTGGATG TCTTAGGTTT TCCATTATCT GTGCCCTCCCT CAGGAAGTCA TTGGATGCTA TTGAACCTGG TTTTAATGTT TTCACCTGCTA
 TCGACCTATT CGAAACCTAC AGAATCCNAG AGTAATPAGA CACGGAGGGA GTCTTCACT AGAATGAGT AACCTGAGT AACTTGGACC AAAATTACAA AAGTGACGNT
 292 L Q I L K P H R L N K M I Q A E R L F D N P D M S G P K L T K V A

2601 CTGGAGTGGT ATTGTTCCAC AGACCTTCCC ATACTCGCC AAACGTACCA GATCCCAATC GCTTCAGNAG CTGTATGGAG TTGCGGTCTA TCTCCCATTTG
 GACCTCACCA TAACAAAGTG TCTGGAAAGG TATGAAAGCG TTGTACTGGT CTAGGGTTAG CGAAGTCTTC GACATACCTC AACGCCAGAT AGAGGGTAAC
 259 V P T T N N W L G E W V E G F Q G S G L R K L L Q I S N R D I E W Q

2701 GTCCACGGTT TTATACGACA NATCAAAATGG AGCTGGGACC TGGATCTTTA AGCATGGTTT CCCCAGCTTG ACACACAGGC CGTCACTTGT CTTGGTGTAG
 CAGGTGCCAA AATATGCTGT TTAGTTTACC TCGACCTGG ACCTAGAAAT TCGTACCAAA GGGTTCGAAC TGTGTGTCG GCAGTGAACA GAACCAATC
 225 D V T K Y S L D F P A P V Q I K L C P K G L K V C L G D S T K T Y

2801 TGGCTCACAA ATTGTTTCAG TGTGAAAG ATTCTCTTTC GGTGAGAAA AAATCCCTCT TCATCCAGTC TTTTAATTT GTAGTGTGTT ACAACTGCTC
 ACCGAGTGGT TAAGCAAGTC ACACTTTTTC TAAGNAGNAG CGCACTCTTT TTTAGGGGGA AGTAGGTGAG AAAATTAGA CATCACAAA TGTGTGACGAG
 192 H S V F E N L T S F I R R R T L F F G G E D L R K I R Y H K V V A

2901 CATCTAAAC TGAAGAGAG AATCTCTCTT TTGCGCTTTC ACTTCTCTG ATTAGAAAG AACCGGTCTT GTTCTCTGAA TATAATAGTT GTTCTCTGCG
 GTAGATTTTG ACTTCTCTC TTAGAGAGGA AAACCGNAG TGAAGAGAGAC TAATCTTTCC TTGGCCAGAA CAAAGACTT ATATTATCAA CAAAGAGACG
 159 G D L V S L S F E G K Q S E S E R I L F S G T K N E S Y L L Q K E A

3001 ATCTGATCTT CCGATTGCTC CAAGAACCA CGGCTCTGCC TGTAGGCTTC TGTCTCAGC CACGTAGTTA GAAGGAATAT AGCCTGTGAG TTGCTGACTG
 TAGACTAGNA GGCTAACGAG GTTCTCTTGGT GCCGAGACGG ACATCCGAG ACAGAGTGG GTGCATCAAT CTCTCTTATA TCGGAACATC AACGACTGAC
 125 D S R G I A G F F W P E A Q L S R D E A V Y N S P I Y G Q L Q Q S

3101 GAGCCATCTC GTCTTTTCTC CAAGTGTCTG GCAAAACACC AGCCCTCATG CAAAGTGTCC AGAATGAA GTTGTGTCAC TGTCTGGAG CTCAAGTCTT
 CTCGGTAGAG CAGAAAGAG GTTCACAGAC CGTTTGGTGG TCGGGAGTAC GTTTCACAGG TCTTGAACCT CAAACAGTGG ACGAGCCTTC GAGTTCAGGA
 92 S G D R R K E L H R A F W W G E H L T D L V Q L K D G A R F S L D

FIGURE 5E

3201 CAGCAGTCCG AGCTGGTAA TCACAAGG CCACAAGTA GTGGCCATGC CTCTGTACT GGGGAGAGCA AAGGGCCCT GATTTTCAA TCACGGTTGA
 GTCGTCAGGC TCGGACCAAT AGTTGTTC GGTGTTTCAT CACCGGTACG GAGACACTGA CCCCTCTCGT TTCCCGGGA CCTAAAAGTT AGTCCCACT
 59 E A T R A Q Y D P L A V F Y H G H R Q S Q P S C L A Q P N E I V T S

3301 CTGTCTGCC TCCGTGACA AACAGGGAG ATAGGTTCT AGTACTCC AGAGCCTCTG ACAGATGTTG CTCATTTGCG CTGGTGGG AGAAGAGGAG
 GAACAGACGG AGGCACCTGT TTGTCCCTC TATCCCAAGA TCCATGAGGG TCTCGGAGAC TGCTACAAAC GAGTAAACAG GAAACCACCCC TCTTCTCCTC
 25 K D A E T S L C P L Y P E L Y E W L R Q C I N S M

3401 CAGGGTTCT CCTCTCCC TTAGTCTCTG CGATCCACCT TATCTTCTT TATCTTCTT CACCAGGCAA CTTTGAAGTC AGCACCACCT CACCATACTT CGAGAGGTAT
 GTCCCGNAGA GGGAGAGGG AATCAGAGAC GCTAGGTGGA ATAGAAGGA GTGTCCGTT GAACTTCAG TCGTGGTTGA GTGGTATGAA GCCTCTCAT

3501 GCAAAGTCCC GTTTCAGATC AGTCCAGCAG CTGGGTGCA GCAAGTCTTA CCTGGAGAGA CTTACCGGT TGCTTCTGT GGTGGAGGT GCTACCCCGA
 CGTTTCAGGG CAAAGTCTAG TCAGGTGTC GACCCACGT GTTCAAGAT GAGCTCTCT GAATGGCCGA ACGAAGAGCA CCGACCTCCA CGATGGGGCT

3601 GGCAAACTG AGCAGGAGCT GGGAGCTGC CCCGTGACG AGTGATCCTT CCACAGAAA GNAGAATAGA CGAATTTCTT GGTGTTGTT TTTAATTTTC

3701 GGCTTTATTT AGACAAATAT CTGAGAACAG AATGGTGCCA TCTTGCCCTT TGTCCCAATA AAAAGTTAGC AAGAGGAAGC TACTAACCCC TGGTAAACCC
 CCGAAATAAA TCTGTTTATA GACTCTGTC TTACCACGGT AGACGGRAA ACAGGTTAT TTTTCANTCG TTCTCTTTCG ATGATTTGGG ACCATTTTGG

3801 TCCACGTCTT GCTTTGCGCA GGTGCACTC GAGGATCTT CCATACCTAC CAGTCTGCG CCTGCAGTTC GCGGCCGGA CTCTAGAGTC GACCTGCAGA
 AGGTGCAGAA CGAAGCGGT CCCAGCTGAG CTCCCTAGNA GGTATGGATG GTCNAGACGC GACGTCCAG CGCCGCGGCT GAGATCTCAG CTGGACGTCT

3901 AGCTTGGCG CCATGGCCCA ACTGTTTAT TGCAGTTAT AATGGTTACA AATAAGCAA TAGCATCACA AATTCACAA ATAAAGCATT TTTTTCACATG
 TCGAACCGC GGTACCGGT TGMCAATA ACGTGGAATA TTACCAATGT TTATTTCGTT TTAAGTGT TATTTCGTAA AAAAGTGAC

19/35

FIGURE 5F

4001 CATTC TAGTT GTGGTTTGTC CAAACTCATC AATGTATCTT ATCATGTCTG GATCGGGAAAT TAATTCGGCG CAGCACCATG GCCTGAATA ACCTCTGAAA
 GTAAGATCAA CACCAACAG GTTTGAGTAG TTACATAGNA TAGTACAGAC CTAGCCCTTA ATTAAGCCGC GTCGTGGTAC CGGACTTTAT TGGAGACTTT

4101 GAGGAACTTG GTTAGGTACC TTCTGAGGCG GAAAGAACCA GCTGTGGAAT GTGTGTCAAT TAGGGTGTG AAAGTCCCCA GGCTCCCCAG CAGGCAGAAG
 CTCCTTGAAC CATCCATGG AAGACTCCGC CTTTCTTGGT CGACACCTTA CACACATCA ATCCACACC TTTCAGGGGT CCGAGGGGTC GTCCGTCTTC

4201 TATGCAAGC ATGATCTCA ATTATCAGC AACCAAGTGT GGAAGTCCC CAGGCTCCC AGCAGGCAGA AGTATGCATA GCATGCATCT CAATTAGTCA
 ATACGTTTC TACGTAGAT TAATCAGTC TTGGTCCACA CCTTTCAGGG GTCCGAGGG TCGTCCGTCT TCATACGTTT CGTACGTAGA GTTAATCAGT

4301 GCAACCATAG TCCCGCCCT AACTCCGCC ATCCCGCCC TAACTCGCC CAGTTCGCC CATTCGCCG CCCATGGCTG ACTAATTTT TTTATTTATG
 CGTTGGTATC AGGCGGGGA TTGAGGCGG TAGGGCGGG ATTGAGCGG GTCAAGCGG GTAAGAGCG GGTACCAG TGAATTAATAA AATAAATATC

4401 CAGAGGCCGA GGCGGCTCG GCCTCTGAGC TATTCAGAA GTAGTGAGGA GGCTTTTGT GAGGCTTAG GCTTTGCAAA AAGCTGTTAA CAGCTTGGCA
 GTCTCCGGCT CCGCGGAGC CGGAGACTCG ATAAGTCTT CATCACTCTT CGAATAAAC CTCGGATCC GAAAACGTTT TTCGACNAIT GTCGMACCGT

4501 CTGGCCGTCG TTTTACAACG TCGTGACTCG GAAACCCCTG GCGTTACCCA ACTTAATCGC CTTCGAGCAG ATCCCCCTTT CGCCAGCTGG CGTAATAGCG
 GACCGGCAGC AATATGTTGC AGCACTGACC CTTTGGGAC CGCAATGGGT TGAATTAGCG GAACGTCTGT TAGGGGAAA GCGGTCCACC GCATATTCGT

4601 AAGAGGCCCG CACCGATCG CCTTCCCAAC AGTTGCGCAG CCTGAATGCG GAATGGCGC TGAATCGGTA TTTCTCTCTT ACGCATCTGT GCGGTATTTT
 TTTCTCCGGC GTGGCTAGCG GGAAGGTTG TCAACGGCTC GGACTTACCG CTTACCGCG ACTACGCCAT AAAAGAGGAA TCGGTAGACA CGCCATAAAG

4701 ACACCGGATA CGTCAAGCA ACCATAGTAC GCGCCCTGTA GCGGCGCATT AAGCGCGCG GGTGTGTG TTACGCGCAG CGTGACCGCT ACCTTGCCA
 TGTGGGTAT GCAGTTTCTGT TGGTATCATG CCGGAGACAT CGCCGCGTAA TTCGCGCCG CCACACCACC AATCGCGCTC GCACGTGGCA TGTGAACGGT

FIGURE 5G

4801 GCGCCCTAGC GCGCGCTCCT TTGCGTTTCT TCCTTTCCTT TCTCGCCACG TTGCGCGGCT TTCCCGGTCA AGCTCTAAAT CGGGGGCTCC CTTTAGGGTT
 CGCGGGATCG CGGGCGAGGA AAGCGAAGA AGCGNAGGA AGAGCGGTGC AAGCGGCCGA AAGGGCAGT TCGAGATTTA GCGCCCGAGG GAAATCCCAA

4901 CCGATTTAGT GCTTTACGGC ACCTCGACCC CAAANACCTT GATTGGGTG ATGTTTCACG TAGTGGGCCA TCGCCCTGAT AGACGGTTTT TCGCCCTTTG
 GGCTAAATCA CGAATGCGG TGGAGCTGG GTTTTGTGA CTAACCCAC TACCAGTGC ATCACCCTGT AGCGGACTA TCTGCCNAA AGCGGGAAC

5001 ACGTTGGAGT CCACGTTCTT TAATAGTGA CTCTTGTTC AAATGGAAC AACACTCAAC CCTATCTCG GCTATCTTT TGAATTAATA GGAATTTGC
 TGCACCTCA GGTGCAAGAA ATTATCACCT GAGACNAGG TTTGACCTTG GTGTGAGTTG GATAGAGCC CGATAAGAA ACTAANTATT CCTNAAACG

5101 CGATTTCCGC CTATTGGTTA AAAATGAGC TGATTTAACA AAATTTAAC GCGAATTTTA ACANATATT AACGTTTACA ATTTATGUT GCACCTCTCAG
 GCTAAAGCCG GATAACCAAT TTTTACTCG ACTAATTTGT TTTTAAATTG CGCTTAAAT TGTTTTATNA TGCANATGT TAAANTACCA CGTGAAGTC

5201 TACNATCTG TCTGATGCCG CATAGTTAAG CCAGCCCCGA CACCGGCCAA CACCGCTGA CGCGCTGTC CGGCTTGTG TGTCTCCCG ATCGCTTAC
 ATGTTAGACG AGACTACGC GTATCAATTC GGTGCGGGCT GTGGCGGACT GCGCGGACT GCGCGACAG ACAGAGGCCG TAGCGNATG

5301 AGACNAGCTG TGACCGTCTC CGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG AAACGCGGA GACGAAGGG CCTCGTGATA CGCTATTTT
 TCTGTTGAC ACTGCGAGG GCGCTCGACG TACACAGTCT CCNAAAGTGG CAGTAGTGG TTTGCGCGCT CTGCTTCCC GGAGCACTAT GCGGATAJAA

5401 TATAGGTTAA TGTGATGATA ATATGGTTT CTAGACGTC AGGTGGCACT TTTCGGGGA ATGTGCGCGG AACCCCTATT TGTATTATTT TCTAANTACA
 ATATCCNATT ACAGTACTAT TATTACCAAA GAATCTGCAG TCCACCGTGA AAAGCCCTT TACACGCGC TTGGGATAA ACANATAAAA AGATTTATGT

5501 TTCAATATG TATCCGCTCA TGAGACAATA ACCCTGATA ATGCTTCAT ATATTTGAAA AAGGAGAGT ATGAGTATTC AACATTTCCG TGTGCCCCTT
 AAGTTTATAC ATAGCGAGT ACTCTGTTAT TGGAGCTATT TACGAGTTA TTATACTTT TTCTTCTCA TACTCATAG TTGTAAAGGC ACAGCGGGA

FIGURE 5H

5601 ATTCCCTTTT TTGCGGCATT TTGCTTTCTT GTTTTGTCTC ACCCAGAAAC GCTGGTGAAA GTAAAGATG CTGAAGATCA GTTGGGTGCA CGAGTGGGTT
 TAAAGGAAAA AACGCCGTAA AACGGAAAGG CAATAACGAG TGGGTCTTTG CGACCACCTT CATTTTCTAC GACTTCTAGT CAACCCACGT GCTCACCCNA

5701 ACATCGAACT GATCTCMAC AGCGTAAGA TCCTTGAGAG TTTTCGCCCC GAAGAACGTT TTCCATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC
 TGTAGCTTGA CCTAGAGTTG TCGCCATTCT AGGAACCTCT ANAGCGGG CTTCCTGCAA AAGTTACTA CTCGTGAAA TTTCNAGACG ATACACCGCG

5801 GGTATTATCC GGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC AGAAAJAGCAT
 CCATAATAGG GCATAACTGC GGGCCGTTCT CGTTGAGCCA CGCGGTATG TGATNAGAT CTTACTGAA CAACTCATGA GTGATCAGTG TCTTTTCGTA

5901 CTTAGCGATG GCATGACAGT AAGAGAAATTA TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGCCA ACTTACTTCT GACAACGATC GAGGAGACCGA
 GAATGCCCTAC CTTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGTA CTCACTATTG TGACGCCGT TGAATGAAGA CTGTTGCTAG CCTCCTGGCT

6001 AGGAGCTAAC CGCTTTTGTG CACAACATGG GGGATCATGT AACTCGCTT GATCGTTGG AACCGAGCT GAATGAAGCC ATACCAACG ACAGCGGTGA
 TCCTCGATTG GCGAAAAAC GTGTGTACC CCTAGTACA TTGAGCGGA CTAGCAACCC TTGGCCTCGA CTTACTTGGG TATGTTTGC TGCTCGCACT

6101 CACCACGATG CCTGTAGCAA TGGCAACAC GTTGGGCAA CTATTAACTG GCGAACTACT TACTTAGCT TCCCGGCMAC AATTATAGA CTGATGAG
 GTGGTGCTAC GGACATCGTT ACCGTTGTG CAGCGGTT GATNATTGAC CGCTTGATGA ATGAGATCGA AGGCCCTTG TTAATTATCT GACCTAGCTC

6201 GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG GTTATTGCT GATAAATCTG GAGCGGTGA GCGTGGTCT CCGGTATCA
 CGCCTATTTC ACGTCTCTG TGAAGACGCG AGCCGGMG GCCGACCGAC CAATAACGA CTATTAGAC CTCGGCACT CGCACCCAGA GCGCCATAGT

6301 TTGCAGCACT GGGGCCAGAT GGTAAAGCCCT CCGTATCGT AGTTATCTAC AGGACGGGA GTGAGGCMAC TATGATGA TATGATAGAC AGATCGCTGA
 AAGTCTGTA CCGCGGTCTA CCATTCGGA GGGCATAGCA TCAATAGATG TGCTGCCCT CAGTCCGTTG ATACCTACTT GCTTATCTG TCTAGCGACT

22/35

FIGURE 51

6401 GATAGGTGCC TCACTGATTA AGCATTTGGTA ACTGTGAGAC CAAGTTTACT CATATATACT TTAGATTGAT TAAAACTTC ATTTTAAATT TAAAAAGGATC
 CTATCCACGG AGTGACTAAT TCGTAACCAT TGCAGTCTG GTTCAATGA GTATATATGA AATCTAATAAATTTTGAAG TAAAAATTAA ATTTTCTTAG

6501 TAGGTGAAGA TCCTTTTTGA TAACTCATG ACCAAATCC CTTAACTGA GTTTTCCTC CACTGAGCGT CAGACCCCGT AGAAAAAGATC AAAGGATCTT
 ATCCACTTCT AGGAAAAAAT ATTAGAGTAC TGGTTTTTAGG GAATTGCAT CAAAGGCAAG GTGACTCGCA GTCTGGGGCA TCTTTTCTAG TTTCCTAGAA

6601 CTTGAGATCC TTTTCTCTG CGCGTAATCT GCTGCTTCCA AACAAAAA CCACCGTAC CAGCGTGT GTTTTTCCTG GATCAAGACC TACCACTCT
 GAACCTCTAGG AAAAAAAGAC GCGCATTAGA CGACGAACGT TTGTTTTTTT GTTGGCATG GTGCCACCA AACAAACGGC CTAGTCTCG ATGGTTGAGA

6701 TTTTCCGAAG GTAACCTGGT TCAGCAGAGC GCAGATACCA AATACTGTT TTTCTAGTGA GCCGTAGTTA GGCACCACT TCAAGAACTC TGTAGCACCG
 AAAAGGCTTC CATTGACCGA AGTCGCTCG CGTCTATGTT TTATGACMAG AAGATCAGAT CGGATCAAT CCGTGTGTG AGTTCTTAG ACATCGTGGC

6801 CCTACATACC TCGCTCTGCT AATCCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTGG TGTCTTACCG GGTGGAATC AAGACGATAG TTACCGGATA
 GGATGTATGG AGCGAGACGA TTAGGACAAT GGTCAACCGAC GACGCTCACC GCTATTTCAGC ACAGATGGC CCAACCTGAG TTCTGCTATC AATGGCCTAT

6901 AGCGCGAGCG GTCGGGCTGA ACGGGGGTT CGTGACACA GCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA
 TCCGCGTCCG CAGCCCGACT TGCCCCCCTA GCACGTGTGT CGGCTGGAAC CTGCTTGTCT GATGTGCTATG GATGTGCTAC TCGATACTCT

7001 AAGCGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCCG AACAGGAGAG CGCACGAGGG AGCTTCCAGG GGGAAACGCC
 TTGCGGCTGC GAAGGGCTTC CCTCTTTTCCG CCTGTCCATA GGCAATTCG CGTCCACGCC TTGTCTCTC GCGTCTCTCC TCGAAGGTCC CCTTTTCCG

7101 TGGTATCTTT ATAGTCTGT CGGTTTTCG CACTCTGAC TTGAGCGTGG ATTTTGTGA TGCTGTGAG GGGGGGAG CCTATGAAA AACGCCAGCA
 ACCATAGAAA TATCAGGACA GCCCCAAGCG GTGAGACTG AACTGCGAGC TAAAAACACT ACGAGCAGTC CCCCCCCTC GGATACCTTT TTGCGGTCTG

23/35

FIGURE 5J

7201 ACGCCGCCCTT TTTACGGGTC CTGGCCCTTT GCTGGCCTTT TGCTCACATG TTCTTTCCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCCTT
TGCGCCGGAA AATGCCAAG GACCGGAAA CGACCGAAA ACGAGTGTAC AGAAGGAC GCAATAGGGG ACTAAGACAC CTATTGGCAT AATGGCGGAA

7301 TGAGTGAGCT GATACCGCTC GCCGAGCCG AACGACGGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA GAGCGCCAA TACGCAACCC GCCTCTCCCC
ACTCACTCGA CTATGGCGAG CGCGTCCGCT TTGCTGGCTC GCCTGGCTCA GTCACTCGCT CCTTGGCCTT CTGCGGGTT ATGCGTTTGG CGAGAGGGG

7401 GCGCGTTGGC CGATTCAATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGGC AGTAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCAAT
CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGCTGAC CTTTGGCCCG TCACTCGGT TCGTTAATT ACACTCAATC GAGTGAGTAA

7501 AGGCACCCCA GGCTTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTAGCG GATAACAATT TCACACAGGA AACAGCTATG ACATGATTAC
TCCGTGGGT CCGAATGTG AAATACGAAG GCCGAGCATA CAACACACCT TAACACTCGC CTATTGTTAA AGTGTGCTT TGTGCGATAC TGTACTAATG

7601 GAATTAA
CTTAATT

24/35

1 GCGGCCGCAG AGAAGCAGA GGATGGGGCT TAGCAGCTGG CAGAGCCAGG AGCGGGGAGG TAGCAGAAAG ACCACAAGTA CAAAGAAGTC CTGAACCTTT
CGCCGGCGTC TCTTGGTCT CTTACCCCGA ATCGTCGACC GTCTGGTCC TCGCCCTCC ATCGTCTTC TGGTGTTCAT GTTCTTCAG GACTTGAAG

101 GGTTTTGCTG CTGCAGCCCA TTGAGACTGA CGACATGGAG CACAAGACCC TGAAGATCAC CGACTTTGGC CTGGCCCCGAG AGTGGCACAA AACACACAAA
CCAAAACGAC GACGTCGGGT AACCTCACT GCTGTACCTC GTGTTCTGG ACTTCTAGTG GCTGAACCG GACCGGGCTC TCACCGTGT TTGGTGTGTT

201 ATGAGTGCCG CXGGCACCTA CXCTGGATG GCTCCTGAGG TTATCAAGGC CTCACCTTC TCTAAGGGA GTGACGTCTG GAGTTTGGG GTGCTGCTCT
TACTACGGC G7CCGTGGAT G7GGACCTAC CGAGGACTCC AATAGTTCCG GAGGTGGAAG AGATTCCCGT CACTGCAGAC CTCAAAACCC CACGACGACA

301 GGGAACTGCT GACCGGGGAG XTGCCATACC GTGGCATTGA CTGCCTTGCT GTGGCCTATG GCGTAGCTGT TAACAAGCTC ACACTGCCAT CCATCCACCT
CCCTTGACGA CTGGCCCCCTC TACGGTATGG CACCGTAATC GACGGAACGA CACCGGATAC CGCATCGACA ATTGTTCGAG TGTACGGTA GGTAGGTGGA

401 GGCC
CCGG

FIGURE 6

25/35

FIGURE 7A

1 ATGAGAGCGT TGGGCGGGA GGGGGCGAG CTGGGCGGCG TGGTGGTGGT TTTTGGGAAAG ATATTGAGGA GTATTATAAA TCAAGATGTC GCTGTGATCA
TACTGTGGA ACCGCGCGT GCGCGCGTC GACGGGAGG AGCAATAAAA AAGCGTTAC TATAAAAGCT GATAATGTTT AGTTCTAGAG GAGACACTAGT

101 AGTGTGTTTT AATCAATCAT AAGAACAAATG ATTCAATCAAT GGGGAAGTGA TATATATATG CGATGGTATG AHAATCCGCG GAAGAGCTGTC GGTGTGCGTT
TCACACAAAA TTAGTTAGTA TTCTTTTATG TAATGTATCA GCGTTCAAT ATATATATAG GGTACCATAG TCTTAGGAGC CTTCTGAGC GCACACGAA

201 GAGACCCGAG AGCTCAGGGA CAGTGAGGA AGCTGTGCT GTGGAAAGTC ATGTATCTTC TTCCATCACA CTGCAAGTTC TGTGCGATGC CCCAGGGGAC
CTCTGGGTC TGAGTCCCT GTCAATGCT TCACAGGGA CAGCTTCAGC TACATAGAG AGGTAGTGT GAGCTTCAGC ACCAGCTAGC GGCTCCCTTG

301 ATTTCCTGTC TCTGGGCTTT TAAGCACAGC TCCCTGAAT GTAGGACACA TTTTGAATTA CAAGACAGAG GAGTTGTTTC CATGTCAT TTGAAATGCA
TAAGGACAG AGACCCAGAA ATTGCTGTCG AGGAGCTTAA GGTGTGCTT AAACTAAAT GTTTGCTGTC CTCACACAAAG GTACCGATAA AACTTTTACT

401 CAGAAACCA AGCTGAGGA TACCTACTTT TTATTGAGAG TGAGCTAGC AATTACAAA TATTGTTAC ATGAGTATA AGAAATACC TGTCTTACAC
GTCTTTGCTT TGAACCTTT ATGATGAAA AATAAGCTTC ACTTGATG TCATGCTAT ATACAAATG TCACTCAT TCTTTATG GAGAAATGTC

501 ATTAAAGA GCTTACTTTA GAAATATGGA AAGCAGGAG CCGCTGCTCT GCATATCTGA GAGGCTTCCA GAGCGATCC TGGATGGCT GCTTTGGGAT
TAATTCCTCT GGAATGAAT CTTTTACCT TTTGCTGCTG GGGACAGA GTATAGACT CTGCGAAGT CTCGCTAGG ACCTTACCA CGAAGCGCTA

601 TCACAGGGGG AAGCTGTAA AGAAGAAAT CTAGCTGTTG TTAAAAAGA GCAAAAGTC CTTCATGAAT TATTTGGAC GCACATAGG TGGTGTGCCA
AGTGTCCCG TTGACATTT TCTCTCTTCA GTTGACAGC AATTTTCTCT GCTTTTTCAT TAAGTACTTA ATAAAGCTG CCGTATATCC AGGACAGGCT

701 GAAATGAAT GGGCAGGGA TGACAGAGC TGTTCAGAA ATAGTAAAT CAAGCTGTC AGACCATTT GGCACAAATTA TTTCTTAAAG TAGGGGAGCC
CTTTATTTGA CCGTCCCTT AGTGTGTC AGAAGTGTTA TGTAGATTTA GTTTCAGTAA TGTGTGTTAA GGTGTTAAT AAAGAAATTC ATGCTGCTTG

FIGURE 7B

801 CTTATATGATTA AGGTGCAAAAT CTGTTGATATTA GAACCAATGGA TTGCTGCTGCA CCTGUGAAAT AGAAGAACAA GUACTGAGAG AGGCGAACATA CTTTUGAGATG
 GAATACCTAT TCCACGTTTC GACAACTACA CTGGTAGCT AAGCCGAGT GGAACCTTAA TCTTTTGTTC CTGAGCTCC TCCGTTGAT GAAACTCTAC

901 ATTACCTATT CAAACAACAG AACTATGATA CCGATTCTCT TTGCTTTTCT ATCATATGTC GCAAGAAAGC ACACGAGTA CTACACTTGT TCCCTTTCAA
 TCAATGATAA GTTGTGTC TTGATACTAT GCTTAAGATA AACHAAACA TAGTAATGAC CTTCTTTTTC TGTGCTAT GATGTGAAC AGGAUAAATT

1001 AGCATCCAG TCAATCAGCT TTGGTTACCA TCGTAGAAA GGGATTATTA AATGCTACCA ATTCAAGTGA AGATTATGAA ATTGACCAAT ATGAAGAGTT
 TCGTAGGCTC AGTTAGCTGA ACCAATGCT AGCATCTTTT CCTAAATAT TTACGATGCT TAAGTTACTT TAACTGCTTA TACTTCTCAA

1101 TTGTTTTCT CTCAGTTTA AAGCCTACCC ACAATACAGA TGTACTGGA CTTTCTCTCG AATATCATT CTGTTUAGC AAGGCTCT TGAATACGGA
 AACAAAAGA CAGTCCAAAT TCGGATGCT TGTATGCT ACATGAGCT GGAAGAGAGC TTTTAGTAAA GGAACCTCG TTTTCCCAAG ACTATTGCT

1201 TACAGCATAT CCAAGTTTG CAATCATAG CACCAAGCAG GAGATATAT ATTCCATGCA GAATATGAT ATGCCAAT TACCANAATG TTCAAGCTGT
 ATTTCTGATA GTTTCANAAC GTTAGTATTC GTGCTGCTGCT CTTTATATA TAAGGTAGCT CTTTACTAC TACGGTTAA ATGTTTTTAC AAGTGGACA

1301 ATATAAGAG GAACTCTAA GTCTGCGAG AAGCTTGGC AAGTCAAGC GTCTGTTCT CCGATGATA CCCATTACCA TCTTGACCT GTTAAGATG
 TATAATCTTC CTTTGAAT CAGATGCTG TTGAAAGCCT TTGATGCTG AGAACAAAG GCTACCTAT GGTATATGUT AGAACCTGGA CTTTCTTAC

1401 TTCAAGCAG TCTCCAACT GCACAAAGA GATCAAGAA GAAATCTGTA ATAGAAAGC TAACAGAAAA GTGTTTGGAC AGTGGTGTG GAGCAGTACT
 AAGTCTGCTC AGAGGCTTGA CTTGCTCTCT CTAGTCTCTT CTTGCTCTCT CTTGCTCTCT TATCTTCTG ATTCTCTTTT CACAAAGCTG TCACCCACAG CTGCTCATGA

1501 CTAAACATGA GTGAAGCCAT AAAGGCTTC CTGCTCAAGT GCTGTGATTA CAATTCCTCT GGCACATCTT GTGAGACGAT CTTTTTAAAC TCTCAGGCC
 GATTTTCTACT CACTTCGTTA TTGTCCTAAG GACCACTTCA CACACCTAT GTTAAGGGA CCGGTGAGAA CAGTCTGCTA GGAATTTTG AGAGTCCG

27/35

FIGURE 7C

1601 CTTTCCCTTT CATCCAAAGAC AACATCTCAT TCTATGCCAAC AATTGGTTTT TTTCTCTCTCT TCAATTGCGT TTTAACCCCTG CTAATTTGTC ACAAGTACAA
 GCAAGGGAAA GTAGGTTCTG TTCTTAGATTA ACATACCTTTC TTAAACCAACA ACAGATCTTTC ACTAACAGCA AATTGGGAC GATTAACAG TGTTCATGTT

1701 AAAGCAATTT AGCTATGAAA GGCAGCTACA GATGCTACAG GTACCCCGAT CCTCAGATAA TCGTACTTC-TACGTTUATT TCAUAGAATA TUAATATGAT
 TTTGCTTAAA TCCATACTTT GCTTCGATTT CATTTCATTTT CATTTCCTTA GGAGTCTATT ACTCATCAAG ATGCANCTAN AGTCTCTTAT ACTTATACTA

1801 CTCAAAATGG-AGTTTCCAAG AGAAATTTTATGAGTTTGGGA AGGTACTAGG ATCAGGTGCT TTGGGAAAAG TCAATGAAGG AACAGCTTAT GGAATTAACA
 GAGTTTACCC TCANAAGGTTT TCTTTTAAAT CTCANAACCT TCCATGATCC TACTCCACGA AACCTTTTC ACTACTTGGG TTGTGGAATA CCTTAATCGT

1901 AAACAGGAGT-CTCAATCCAG GTTACCGTCA AATGCTGAA AGAARAGCA GACAGCTCTG AAAGAGAGG-ACTCATGTCA GAACTCAAGA TGAATGACCC-
 TTGTCTCTCA GAGTTAGGTC CAATGGCAGT TTATCCGACTT TCTTTTTCGT CTGTGAGAG TTTCTCTCCG TGAATACAT CTGAGTTCT ACTACTGGGT

2001 TGTGGGAGG-CACGAGATAA TTGTGAACCTTGGTGGGGGGG TGCACACTGT CAGGACCAAT TTACTTGATT TTGTAATACT GTTGTATGG TGAATCTTG TGAATCTCTG
 CCAACCTTCTG GTGCTCTTAT AACACTTGGG CGACCCCGCG ACCTGTGACA GTCTGTGTTA AATGAATTA AACCTTATGA CAACGATACC ACTGAGAGAG

2101 AACATATCTAA-GAAGTAAAG AGAAATTTT CACAGGACTT GACAGAGAT TTCAAGGAA CACAATTTCA GTTTTACCC CACTTTCCAA TCACATCCAA
 TTGATAGATT CTTCAATTTT TCTTTTAAA GTCTCTGAA CCTGCTCTGA AAGTTCTTT GTGTTAAAGT CAAAAATGG GTGAAAGTTT AGTGTAGGTT

2201 ATTCCAGCAT-GCTGGTTCA AGAAGATTG-AGATACACCC GACTCGGAT CAATCTCAG GCTTTCATGG GAATTCATT CACTCTGAAG ATGAATTTGA
 TAAAGTCGTA GGGACCAAGT TCTCTTCAAG TCTATGTGG CTTGAGCTTA GTTTAGAGTC CCGAAGTACC CTTAAGTAAA GTGAGACTTC TACTTTAAT

2301 ATATGAAAAC CAAAAAGGC TGAAGAGAGA GGAGGACTTG AATGTGCTTA AATGTGCTTA TCTTCTTTCG TTTGCTATC AATTTGCCAA AGAATGGA
 TATACTTTG GTTTTTCG ACCTTCTCT CTTCTGAG TTAACAGAA TTACAGCAAT GTAAACTCT AGAAGAAACG AACGTATAG TTCAACGTTT TCTTACTCT

FIGURE 7D

[illegible]

350) ~~AGACATATCAT~~ GAGTGATTCC AACTATGTTG TCAGGGCA. TCCCGTGTG CCTGTAAAT GAGTGGCCC GGAAGCCG TTTGAAGCA TCACACCAT
CTCTATTTGTA CTCTACTTAGG...TTGATACAA ~~TTTTTCCGTT~~ ACGGGCAGAC GGACATTTTA CCTACCGGGG GCTTTGGAC AACCTCCG AGATGTGTA

2601 TAAAGGTGAT-GTCTGTGTCAT-ATGGAATATT-CTGTGGGA-ATCTTCTCAC-TTGGTGTGA-TCCTTACCTT-GGCATTCCG-TTATGCTAA-CTTCTACAAA
ATTCTCACTA CAGACCAGTA TACCTTATTA-TTACACCTT-TAGANATTC-AACACACATT-AGGAATGGG-CCGTARGGC-AACTACGATT-GAAGATGTTT

CTGATTCAAA-ATGGAGTTAA AATGGATCAG CCATTTTATG CTACAGAAAG AATAATACATT ATAAATGCAN¹ CCTGCTGGGC TTTTGACTCA AGGAACAGCG
GACTAAGTTT TACCTAAATT TTACCTAGTC GGTAAATATC GATCTCTT² TTATATATGTA TATTACCTTA GGACGACCCG AAACTGTAGT TCGTTTGCCG

2001 CATCCCTCC-TAAATTGACT 'TCGTTTTTAG: GGCAGATGCA GAAGAAGCA TGATATCAGAA 'TOTGTATGTCG CBTGTTTTCG ANTGCTCTCA
GTAGGAGAGG; ATTAAATCGA AGCAAAATC CTACAGTGA CCGCTACGT CTCTCTTCGCT ACATGCTCTT ACACCTACCG GCACAAGCC TTACAGGAGT

2901 CACCTACCAA,ACAGGGGAC CTTTCACGCG:AGAGATGGAT:TTGGGGCTAC TCTCTCCGCA GGCTCAGGTC.GAAGATTCGT AGAGGAACAA TTTAGTTTAA
GTGGATGGTT,TTGTCCGGTG GAAGATGCTC:TCTCTACCTA,ACCCCGATG AGAGGGGCT CCAGATCCAG CTCTACGA TCTCCTTGTT AATCAAAAT

3001-AGGACCTCAT⁺CCCTCCACCT⁺ATCCCTAACCA⁺GGCTGAGAT⁺TACCAAAACA⁺AGGTTAATTT⁺CATCACTAAA⁺AGAAATCTA⁺TTATCAACTG⁺CTGCTTCACC⁺
-TCCTGAGATGA⁺GGGAGGTTGGA⁺TAGGGATTTGT⁺CCGACATCTA⁺ATGGTTTGT⁺TCCAAATTAAA⁺GTAGTGATTT⁺TCTTTTATGAT⁺ATAAGTTGAC⁺GACGAGATGCG⁺

1101, AGACTTTTCT, CTAGAGAGCG
;TCTGAAGA: FATTCTCGC

29/35

FIGURE 8A

1 TCGGCTCCA CCGGCCAGG GAGATCAGA CTTGKXKGG CAGRGCCC CCAACTCAG TTGGATCCT ACCGAGTGA GGGGGGCA TGGAGTCCG
 1 AGTCGAGOT GGGGGGTCC CTCTAGTCT GAGCCCCCC GTCGGGGG GTTTGATGTC AGGCTAGGA TGGGCTCACT CCGCGCGGT ACCTGAGGC
 M E L R

101 GGTGCTGCTC TGTGGGCTT CGTTGGCCG AGCTTTGGA GAGACCTGC TGAACACAAA ATTGGAACT GCTGATCTGA AGTGGTGAAC ATTCCCTCAG
 CACGACGAG AGCACCGAA GCACCGGCTT TGTAACTT CTCTGGAGG ACTTGTGTTT TAACCTTTGA CGACTAGACT TCACCCACTG TTAGGGAGTC
 5 V L L C W A S L A A A I E E T L L N T K L E T A D L K W V T P P Q

201 GTGGACGGC AGTGGGAGGA ACTGAGGCG CTTGATGAG AACAGCAG CTTGTCACC TACGAGTGT GTGAGTCCA GGTGCCCC GGGCAGGCC
 CACCTGCCG TCACCTCCT TGACTGCGG GACTACTCC TTGCTGTGTC GCACCGGTG ATGCTTACA CACTGCAGT CGACCGGGC CCGTCCCGG
 39 V D G Q W E E L S G I D E E G A V H V Y A T L R F T H L E C L S L P R A Q

301 ACTGCTTGG CACAGTTGG GTCCACGCG GRRGRRCTT CCAGRTTAC GCGAGCTGC GTTCACCAT GTTCGATGC CTGCTCTGC CTGGGCTGC
 TGACCGAGC GTGTCCACC CAGGTTCG CCGCGGCA GTTGACATG CGTGGCAG CGAGTGTGA CGAGCTCAG GACAGGAGG GAGCCGACC
 72 W L R T G W V P R R G A V H V Y A T L R F T H L E C L S L P R A Q

401 GCGCTCTGC AAGGAGCCT TCACCTGCTT CTAATATGAG AGCGATGCG ACACGGCCAC GGGCTCTACG CCAAGCTTGA TGGAGACCC CTACATCAG
 CCGAGGAGG TTCTCTTGA AGTGGCAGA GATGATATCT TCGCTAGCC TGTGCGGTG CCGGAGTGC GTTCGAGCTT ACCTCTTGG GATGTAGTTC
 105 R S C K E T P T V F Y Y E S D A D T A T A L T P A W H E N P Y I K

501 GTGGACAGG TGGCGCGGA GCATCTCACC CGGAGCGCC CTGGGCGGA GGGCAGCGG AGGTGAATG TCAAGACCTT GGTCTTGGGA CCGCTCAGCA
 CACCTGTGC ACCGGGCTT CGTAGAGTGG GCTTTCGCG GACCCGCTT CCGGTGCCC TTCCACTTAC AGTTCTTGA CCGAGACCTT GCGGAGTGT
 139 V D T V A A E H L T R K R P G A E A T G K V N V K T L R L G P L S K

601 AGGTGCTCTT CTACTGCGC TTCCAGGACC AGGTGCGCT CATGGCGCT CTATCCTGC ACCTTCTTA CAAGAATGC GGGCAGCTGA CTGTGAACT
 TCCGACCGA GATGACCGG AGGTCTCTG TCCACGAC GTACCGGAC GATAGGAGG TGGAGAGAT GTTTTACG CCGTCCACT GACACTTGA
 172 A G F Y L A F Q D Q G A C H A L I S L H I P Y K K C A Q L T V N L

30/35

FIGURE 8B

801 TACTGCGGTG AGGATGGCCA GTGGCCGGA CAGCGGTCA CGGGCTGAG CTGTGCTCCG GGTTCGAGG CAGCTGAGGG GAAACACCAAG TCCCGAGCCT
 ATGACGGCAC TCTTACCGGT CACCCGCTT GTTGGCCACT GCTGGAGCTC GACACGAGC CCCAGCTCC GTCACTCC CTTGTGGTTC AGGCTCGGA
 218 Y C R E D G Q W A E Q P V T G C S C A P G P E A A E G N T K C R A C

901 GTGCCAGGG CACCTTCAG CCCCTTTCAG GAGAGGTTC CTCGAGCCA TCCCGACCA ATAGCCACTC TACACCAAT GATCAGCG TCTGCCAGT
 CACGGTCCC GTGAGTTT GGGAGTTT CTTTCCAG GACGGTGGT ACGGTGGT TATCGGTGAG ATTGTGTAA CCTAGTGGC AGAGGTTCAC
 272 A Q G T P K P L S G E G S C Q P C P A N S H S N T I G S A V C Q C

1001 CGCGTCCGG TACTTCCGG CAGGACAGA CCCCGRGT GCACCTGCA CCAGCCCTCC TTGGCTCCG CGAGCGTGG TTTCGGCT GAACTGCTCC
 GCGCAGCCC ATGAGGCC GTGGTGTCT GGGGCCCA CTTGGGACT GTGGGGAGG AACCGAGGC GCCTGCACC AAGCGCGA CTTGCCGAGG
 305 R V G Y F R A R T D P R G A P C T T P P S A P R S V V S R L N G S

1101 TCCCTGCACC TGGATGGAG TCCGCCCTG GAGTCTGGT GCCAGAGA CTCACCTAC GCCTCCCT GCGGAGTGG CCGACCCGA GCTCTCTTG
 AGGAGCTGG ACCTTACCTC ACGGGGAC CTCAGACCAC CGCTCTCTT GGAGTGGATG CGGAGCGGA CGCCCTCAC GGTGGGCT CCGAGGACAC
 338 S L H L E W S A P L F S G G R E D L T Y A L R C R E C R P G G S C A

1201 CGCCCTGCG GAGAGACTG ACTTTGACC CCGGCCCG GACCTGGTG GAGCCCTGG GAGCTAGTG AGGCTAGCT CCTGACTTCA CCTATACCTT
 GCGGAGGCC CCTCTGAG TGAATGAG GATCGGTGC CCGCGGCG CCTGACCAC CTCGGACCC ACCACNAGC TCCCGATGCA GACTGAGT GATATGGA
 372 P C G G D L T P D P G P R D L V R P W V V V R G L R P D P T Y T F

1301 TGAGTCACT GCATTGAG GGTATCTTC CTTAGCAG GGGCCGTC CATTTGAGC TGTCATGTC ACCACTGACC GAGAGGTACC TCTGCGAGT
 ACTCCAGTGA CGTAACTTGC CCATAGGAG GATCGGTGC CCGGGGAG GTAACTCG GATTTACAG TGGTACTCG CTCTCCATGG AGGAGTCAC
 405 E V T A L N G V S S L A T G P V P F E P V N V T T D R E V P P A V

1401 TCTGACATCC GGTGAGCG GTCTCAACC ATCAGTTTA GCCTGCTTG GCTGTCTCC CGGACACCA GTGGGGCTGT GCTGACTAC GAGGTCAAT
 AGACTGTAG CCCACTGCG CAGGAGTGG TCGTCAACT CGGACGGAC CCGACGAG GCGGTGGT CACCCGACA GAGCTGATG CTCCAGTTA
 438 S D I R V T R S S P S S L S L A W A V P R A P S G A V L D Y E V K Y

FIGURE 8C

1501 ACATGATGAA GGGGGCTTGG GTTCCAGCA GGTGGGTTT CTTGAGAGC TCAGAAACC GGGCAGACT GCGGGGCTG AAGCGGGAG CCAAGTACTT
 TGTACTCTT CCGCGGCTC CAGGGTGT CTTACTGTA GACTTCTG AGTCTTTGG CCGTCTGA GCGCGGAC TTGCGGCTC GGTGATGTA
 472 II E K G A E G P S N V R F L K T S E N R A E L R G L K R G A S Y L

1601 GGTGAGTA GGGGGGCT CTTAGGCGTG CTAGGCGG CTAGGCGG TTGCGGAG ACATGACG CAGAGCCAA CTGATGAG GCGAGGCTG GCGGAGCAG
 CCACGTCCAT GCGCGGCGA GACTCGGCG GATTCGCG AAGCGGTC TTGTAGTGC GTCTGGTT GACTACTCT CCGTCCGAC CCGCTCTGC
 505 V Q V R A R S E A G Y G P F G Q E II S Q T Q L D E S E G W R E Q

1701 CTGGGCTGA TTGCGGCAC GCGAGTCTG GGTGCTGCTT TGTCTCTGT GGTCTTGT GTCGACTTC TCTGCTGAG GAGCAGAGC ATGTGTAGAG
 GACCGGACT AACCGCGTG CCGTCAGCAC CACACCGAG ACCAGACCA CCACTAAC CAGGTCAAG AGACGAGTC CTCTCTCTC TTACCTCTC
 538 L A L I A G T A V V G V V L V L V V L V A V L C L R K Q S N G R E

1801 AAGCAGATA TTGGACAA CAGGACAGT ATCTATCG ACATGACT AAGTCTACA TCGACGCTT CACTTATGA GACCTTAATG AGCTGTGAG
 TTGTCTTAT AAGCTGTT GTGCTGTA TAGAGTAGG GTTACCTGA TTCCAGTGT AGCTGGGAA GTGATACTT CTGGATATC TCGACACTC
 572 A E Y S D K H G Q Y L I G H G T K V Y I D P F T Y E D P N E A V R

1901 GGAATTTGA AAGAGATCG ATGTCTCTA GGTCAAGAT GAAGAGTGA TTGTGACG TTAGTTTGG GAGTGTG GCGGCGGCT CAAGGCGCA
 CTTAAAGT TTCTCTAG TACAGAGGAT GCACTCTTA CTTCTCACT AACCATCC ACTCAACCG CTCACACCG CCGCGCGCA GTTCCGCGT
 605 E F A K E I D V S Y V K I E E V I G A G E F Q E V C R G R L K A P

2001 GGAAGAAAG AGAGCTGTGT GGCATCAAG ACCGTGAAG GTTCTACG GTTGTGCG GCGGCGG CCGCTGAGT TCTGAGCA GCGCTGCTC ATGCGGCACT
 CCTTCTTCC TCTGACACA CGTTAGTTC TGGGACTTC CACCATGTC CCGCGCTC GCGCACTCA AGACTCGCT CCGAGGTAG TACCGGCTCA
 638 G K K E S C V A I K T L K G G Y T E R Q R R E F L S E A S I H G Q P

2101 TCGAGCACC CAATATCATC CCGCTGAG GGTGTGAC CACAGCAT CCGCTCATG TTCTCAGCA GTTGTGAG AACGCGGCTC TCGACTCTT
 AGCTCTGG GTTATAGTAG GCGGACTTC CCGACAGTG GTTGTCTAC GCGCACTACT AGAGTGTCT CAAGTACTC TTGCGCGG ACCTGAGGA
 672 F II P N I I R L E G V V T N S M P V M I L T E F M E N G A L D S F

FIGURE 8D

2201 CTTGCGGCTA AACGACGAC AGTTACAGT CATCAGCTC GTGGCATGC TGGGGGAT TGGCGGAC ATGCGTACC TTGCGAGAT GAGTACATC
 GGAGCGCAT TTGCTGCTG TCAGTGTCA GTAGTGTGAG CACCGTAGC AGCGCCGTA GCGAGCCG TAGCCATGG AACGCTCTA CTCGATGAG
 705 L R L N D G Q F T V I Q L V G M L R G I A S G M R Y L A E M S Y V

 2301 CACGAGACC TGGCTGCTG CAACATCTA GTCAACATCA ACCTGCTG CAAATGTCT GACTTTGGC TTTCGCAAT CCGAGAGAG AACCTTCTG
 GTGCTCTG ACCGAGAC GTTGTAGAT CAGTTTCTGT TGGAGCAGC GTTTCACAGA CTGAACCGG AAGGCTAA GAGCTCTC TTGAGAGGC
 738 H R D L A A R N I L V N S N L V C K V S D F G L S R F L E E N S S D

 2401 ATCCACCTA CACGAGCTC CTGGAGGAA AGATTCCAT CCGATGGAT GCGCGGAG CCAATGCTT CCGAAGTTC ACTTCGCA GTGATGCTG
 TAGATGGAT GTGCTGAGG GAGCTCTT TCTAAGGTA GGCTACCTGA CCGGCTCC GTTAACGAA GCGCTCAAG TGAAGCGT CACTACGAG
 772 P T Y T S S L G Q K I P I R W T A P E A I A F R K F T S A S D A W

 2501 GAGTTACGG ATTGTGATG GAGAGTGT GTCATTTGG GAGAGCGGT ACTGGGAT GAGCAATCAG GAGTGTATCA ATGCCATTA ACAGACTAC
 CTCATGCCC TAACACTACA CCTCCACTA CAGTAACCC CTCCTCGCA TCACCTGTA CTCGTATGTC CTGCACTAGT TAGGTAACT TGTCTGTATG
 805 S Y G I V H W E V M S F G E R P Y W D M S N Q D V I N A I E Q D Y

 2601 CGGCTGCCCC CGCCCCAGA CTGTCCACC TCCCTCCACC AGCTATGCT GAGTGTGG CAGAAAGACC GGAATGCCG GCGCTGCTTC CCGGAGTGG
 GCGGACGGG GCGGGTCT GACAGGTGG AGGAGGTGG TCGATGAGA CCTGACACC GTCTTTCTGG CCTTACGGC CCGGCGAAG GCGGTCCACC
 838 R L P P P P D C P T S L H Q L M L D C W Q K D R N A R P R F P Q V V

 2701 TCAGCGCCT GAGACAGATG ATCGGAAACC CCGCAGCCT CAATCTG GCGCGGAGA ATGCGGCGC CTCACACCT CTCCTGACC AGCGGAGCC
 AGTCGCGGA CCTGTTCTAC TAGGCTTGG GCGGTGGA GTTTAGCAC CCGGCTCT TAGCGCCCG GAGTGTGGA GAGGACCTG TCGCGCTCG
 872 S A L D K H I R N P A S L K I V A R E N G O A S H P L L D Q R Q P

 2801 TCACTACTA GCTTTTGGT CTGTGGGGA GTGGCTTGG GCCATCAAA TGGAGATA CGAAGAAAT TTGCGAGCG CTGGCTTGG CTCCTTGGAG
 AGTGATGAGT CGAAACCGA GACACCGCT CACCGAGCC CGGTATTTT ACCCTCTAT GCTTCTTCA AGCGTGGC GACCGAACC GAGGAGCTC
 905 H Y S A F G S V G E W L R A I K M G R Y E S P A A A G F O S F E

FIGURE 8E

2901 CTGGTCAGCC AGATCTCTGC TTAGGACCTG CTCGAAATCG GATCACTCT GTGGGGACAC CAGAGAAA TCTTGGCCAG TGTCCAGCAC ATGAGTCCC
 GACCACTCGG TCTAGAGACG ACTCTTAAAC GAGGTTTAC CTCAGTGAAGA CCGCCCTGTG GTCTCTTTT AGAACCGTC ACAGTCTGT TACTTCAGGG
 938 L V S Q I S A E D I, L R I G V T L A G H Q K K I L A S V Q H M K S Q

3001 AGGCCAAGCC GGAACCCCG GTTGGGACAG GAGGACCGC CCGCAGTAC TACCTCGAG GAATCCCA CCGCAGGAC ACCGCTCC CATTTCGCG
 TCGGTTCCG CCGTTGGGCG CCACCTGTG CTCCTGGCG GCGCTCATG ACTGAGCTC CTGAGCGGT GCGTCCCTG TCGCGAGCG GTAAAGGCC
 972 A K P G T P G G T G G P A P Q Y O P A G T P H P R D T A S P F S G

3101 GGCAGAGTGG GGACTCACAG AGGCCCCAG CCCTGTGCC CGCTGGATTG CACTTTGAG CCGTGGGTG AGGAGTTGC ATTTGGAGA GACAGGATTT
 CCGTCTACC CCGAGTGTG TCCGGGGTC GGGACACGG GGCACCTAAC GTTAAACTG GGCACCCAC TCCTCAAGC TTAACTCT CTGTCTAAA
 1005 A E W G L T E A P S P V P R W I A L O A R G V R S W Q F G E T G F

3201 GGGGTTCTG CCATATATAG AGGGGAAAT CACCCCCAG CCACCTCGG GAACTCAGA CCAAGGTGA GGGGCTTT CCCTCAGGAC TGGGTGAC
 CCCCAGAC GGTATTATCC TCCCTTTTA GTGGGGTC GGTGAGGCC CTTAGGTCT GTTCCACT CCGCGGAAA GGGAGTCTG ACCCACATG
 1038 G G S A I I G G E N H P P A T S G N S R P R V R A P F P Q D W V O P

3301 CAGAGGAAA GGAAGTCCC AACATCTCC AGCTCCCA GTGCCCCG TCACCTTGT GGTGCTTC CCGCAGACCA AAGAGAGTGT GACTCCCTTG
 GTCTCTTTT CTTACAGGG TTGTAGAGG TCGAGGGGT CACGAGGG AGTGAACTA CCGACCAAG GCGCTCTGT TTCCTCACA CTGAGGAGC
 1072 E E K E V P N I S Q P P Q V P P S P O W V R S R R P K R V O L P C

3401 CCAGTCCAG AGTGGGGGG CTGTCCACAG GGGCAAGAG GGTGTACG GCCCAGTAC AAATCATG GGGTTGTAG TCCCACTTG CTCTGTAC
 GGTGAGGTG TCACCCCGCC GACAGGGTCC CCGTTCTTC CCCACATCC CGGTCACTG TTTTGTAC CCGCAACATC AGGTTTGAAC GACGACAGT
 1105 Q L Q S G G A V P G G K K G C Q G P V T K S L G F V V P T C C H

3501 CACCAACTC AATCATTTT TTCCCTTGA AATGCCCTC CCGCAGTGC TGCCTCATA TTGAGGTTT TTGAGTTTG TTTTGGTCT TAATTTTCT
 GTGGTTTGA TTAGTAAAA AAGGAAACAT TTACGGGAG GGGTGGACG ACGAAGTAT AACTCCAAA AACTCAAAAC AAAAACGAGA ATTTAAAGA
 1138 H Q T Q S P F S L V N A P P P A A A P I L K V F E F C F W S O F F S

34/35

FIGURE 8F

3601 CCGGTGCCC TTTTGTTC TCGTTTGT TTTCTACCG TCTTGTGAT AACTTGTGT TGGAGGAGAC CTGTTTCACT ATGCGCTGCT TTGCCCAAGT
 GGGCAAGGG AAAAAAAG AAGCAAAACA AAGATATGC AGGACAGTA TTGAACACA ACCTGCTTGT GACAAAGTGA TACCGAGGA AAGGTTCA
 1172 P F P P C F P V L F P V L F P Y R P C H N P V L E G T C P T H A S F A Q V

3701 TTAACAGGG GCCATCATC ATGTCTTTT CCAAGACAGT GCTTGTGCA TCCACATCC CCGACCCCG CCGGAGACC CCAAGCTGTG TCCTATGAAG
 ACTTGTGCCC CGGTAGTAG TACAGACAA GGTCTTGTCA CGGACCAAGT AGGTTGTAGG GGCCTGGGC GACCCCTGGG GATTCCACAC AGGATACTTC
 1205 E T G A H H H V C F Q N S A L V I P H P R T P P G T P K L C P H K

3801 GGGTGTGGG TGAGGTAGT AAGGGCGG TAGTGTGTG TGAACCCAG AACGGAGCG CGGTGCTTG AGGCTTCTT AATTATATT TAAAAAGTA
 CCCACACCCC ACTCCATCAC TTTTCCGCC ATCAACCACC ACCTGTGTC TTGCTGTGG GCCACGACC TCCCAAGAA TTTAATATAA ATTTTTCAT
 1238 G C G V R O O K G R O L V V E P R N G R R C L E G F L N Y I O K S N

3901 ACTTTTGTG TAAATAAAG AATATGGAC GTGTCCGAG TCCAGGGTA AAAAAAAA AAAAAAAA
 TGAATACAT ATTAATTTT TTTACCTG CACAGGTG AGGTCCCAT TTTTTTTTT TTTTTTTTT
 1272 F L Y K O K K H G R V P A P G V K K K K K K

- FIGURE 9
ARNILVNSNLVCKVSDFGLSRFLDDTSDPTTYSALGGKIPMRWTAPEAIQYRKFSAS
- FIGURE 10
NVLKSPNHVKITDFGLARLLEGEDEKEYNADGGKMPIKWNALCEIHRYRKFTTHQS
- FIGURE 11
NCHLAGDMTVCVADFGLSWKIYSGATIVRGCASKLPVKWLAGSLADNLYTVHS
- FIGURE 12
NCLVGKNYTIKIADFGMSRNLYSGDY
- FIGURE 13
TRNILENENRVKIGDFGLTKVLPQDKYYKVKEPGESPIFWYAPESLTESTLPSVASD
- FIGURE 14
ARNILVNSNLVCKVSDFGMSRVLEDDPEAAAYTRGGKIPIRWTAPEAIYRKFTSASD

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 93/00586

I. CLASSIFICATION F SUBJECT MATTER (If several classification symbols apply, indicate all) ⁶		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. 5 C12N15/12; C12N15/54; C12N9/12; //C12Q1/68, C12N15/11		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
Int.Cl. 5	C12N ; C12Q ; C07K	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁸		
III. DOCUMENTS CONSIDERED TO BE RELEVANT⁹		
Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
A	<p>NEURON vol. 6, no. 5, May 1991, pages 691 - 704 LAI, C. & LEMKE, G. 'An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system' see the whole document</p> <p style="text-align: center;">--- -/-</p>	1-7
<p>¹⁰ Special categories of cited documents : ¹⁰</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report
07 JUNE 1993		02 -07- 1993
International Searching Authority		Signature of Authorized Officer
EUR PEAN PATENT OFFICE		ANDRES S.M.

III. DOCUMENTS CONSIDERED TO BE RELEVANT

(CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 88, no. 23, 1 December 1991, WASHINGTON US pages 10411 - 10415 HOLTRICH, U. ET AL. 'Two additional protein-tyrosine kinases expressed in human lung: fourth member of the fibroblast growth factor receptor family and an intracellular protein-tyrosine kinase'</p>	1-7, 13-15
A	<p>GENE vol. 110, no. 2, 15 January 1992, AMSTERDAM NL pages 205 - 211 BRAUNINGER, A. ET AL. 'Isolation and characterization of a human gene that encodes a new subclass of protein tyrosine kinases' see the whole document</p>	1-7, 13-15
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 87, November 1990, WASHINGTON US pages 8913 - 8917 PARTANEN, J. ET AL. 'Putative tyrosine kinases expressed in K-562 human leukemia cells' see especially clone JTK10</p>	1-7
A	<p>MOLECULAR AND CELLULAR BIOLOGY vol. 9, no. 4, April 1989, WASHINGTON US pages 1587 - 1593 HAO, Q.-L. ET AL. 'Isolation and sequence analysis of a novel human tyrosine kinase gene' see the whole document</p>	1-7, 16-18
P,A	<p>WO,A,9 214 748 (AMERICAN CYANAMID COMPANY) 3 September 1992 see the whole document</p>	1-12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 93/00586

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
obscurities.
see additional page
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

OBSCURITIES

Claims 2,3,5 and 7 as filed in the application are in contradiction with the description. Indeed, they are dependent on claims 1,4 or 6 which are related to DNAs or proteins of human MEGAKARYOCYTIC origin, whereas they claim some sequences which are not from such an origin, if taking reference to page 8 of the description. Therefore the search of these claims has been made independently of claims 1,4 and 6.

Claims 14 and 15 relating to the nucleotide and the protein sequences of SAL-D4 are filed as being dependent on claim 10 which relates to SAL-S1. They have been interpreted and searched as depending on claim 13.

Claims 17 and 18 relating to the nucleotide and the protein sequences of LpTK 3 are filed as being dependent on claim 14 which relates to SAL-D4. They have been interpreted and searched as depending on claim 16.

Claims 20 and 21 relating to a DNA expression vector and a cell transformed therewith, are filed as being dependent on claim 17 which relates to a protein (LpTK 3). They have been searched as depending on claim 19.

Finally, claim 2, point a) relating to SAL-S1 refers erroneously to SEQ ID 7 which describes SAL-D4. It has been interpreted as referring on SEQ ID 5 and 17).

US 9300586
SA 69794

07/06/93

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9214748	03-09-92	EP-A- 0536350	14-04-93